

GenCore version 5.1.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 16:34:19 ; Search time 6488 Seconds
(without alignments)
16982.618 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaattacatt.....tgctctgttcattagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_da.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3786	100.0	3786	8	YSAALS1	L25902 Candida alb
2	1834.4	48.5	3360	8	CAU87956	U87956 Candida alb
3	1708.6	45.1	4383	8	AF051313	AF051313 Candida a
4	1526.4	40.3	4569	8	AF272027	AF272027 Candida a
5	1392.8	36.8	4723	8	AF025429	AF025429 Candida a
6	1188.8	31.4	3813	8	AF068866	AF068866 Candida a
7	1069.4	28.2	1071	6	AR044076	AR044076 Sequence
8	1069.4	28.2	1071	6	I65632	I65632 Sequence 1
9	1016.6	26.9	1404	8	CANALS2S1	AF024582 Candida a
10	1006	26.6	1404	8	CAIALS1	AF024580 Candida a
11	905.8	23.9	4332	8	AF075293	AF075293 Candida a
12	874.4	23.1	6897	8	AF201684	AF201684 Candida a
13	842.8	22.3	1407	8	CANALS4S1	AF024584 Candida a
14	836.4	22.1	1407	8	CANALS4S1	AF024586 Candida a
15	805.4	21.3	1404	8	AF229989S1	AF229989 Candida a
16	731.8	19.3	1007	8	AF202529	AF202529 Candida d
17	507	13.4	995	8	AF202530	AF202530 Candida d
18	463.6	12.2	1013	8	AF201685	AF201685 Candida d
19	394	10.4	1299	8	AF075294	AF075294 Candida a
20	362.2	9.6	1058	8	AF201686	AF201686 Candida t
21	348.2	9.2	425	6	AR044080	AR044080 Sequence
22	348.2	9.2	425	6	I65636	I65636 Sequence 5
23	331.2	8.7	680	8	AF189016	AF189016 Candida a
24	202.4	5.3	1400	8	AB002099	AB002099 Candida t
25	198.6	5.2	13370	8	SPAPB2C8	AL590602 S.pombe c
26	170.8	4.5	172307	2	AC044842	AC044842 Homo sapi
27	141.6	3.7	377	8	AF211866	AF211866 Candida t
28	135.8	3.6	132449	9	AL365272	AL365272 Human DNA
29	134.4	3.5	20325	8	SPCPB16A4	AL591677 S.pombe c
30	132	3.5	380	8	AF211865	AF211865 Candida t
31	130.4	3.4	158615	2	AC117835	AC117835 Rattus no
32	126	3.3	185994	2	AC002042	AC002042 Homo sapi
33	125.8	3.3	175748	2	AC120669	AC120669 Rattus no
34	125.8	3.3	180903	2	AC125859	AC125859 Rattus no
35	121.6	3.2	12151	1	U23947	U23947 Mycoplasma
36	116	3.1	53352	5	AL592077	AL592077 Zebrafish
37	112.2	3.0	157	8	AF035757	AF035757 Candida a
38	111.8	3.0	180668	2	AC020857	AC020857 Mus muscu
39	111.6	2.9	666	8	AF413050S2	AF413051 Zea mays
40	111.2	2.9	22398	5	FRU271723	AJ271723 Fugu rubr
41	110.2	2.9	141017	2	AC116962	AC116962 Dictyoste
42	109.8	2.9	35412	8	SPCC188	AL049662 S.pombe c
43	109.8	2.9	38141	8	SPAC8A4	266569 S.pombe chr
44	109.6	2.9	127354	2	AC117014	AC117014 Rattus no
45	109.6	2.9	155019	2	AC117361	AC117361 Rattus no

ALIGNMENTS

RESULT 1	YSAALS1	3786 bp	DNA	linear	PLN 03-MAY-2000
LOCUS	Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.				
DEFINITION	Candida albicans				
ACCESSION	L25902				
VERSION	L25902.1				
KEYWORDS	GI:704426				
SOURCE	Candida albicans				
ORGANISM	Candida albicans				
REFERENCE	1 (bases 1 to 3786)				
AUTHORS	Hoyer, L.L., Scherer, S., Shatzman, A.R. and Llivi, G.P.				
TITLE	Candida albicans ALS1: domains related to a Saccharomyces				

cerevisiae sexual agglutinin separated by a repeating motif
Mol. Microbiol. 15 (1), 39-54 (1995)
95272392
PUBMED
7752895

FEATURES
source
1..3786
/organism="Candida albicans"
/strain="B792"
/db_xref="taxon:5476"
1..3783
/gene="ALS1"
1..3783
/gene="ALS1"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC41649.2"
/db_xref="GI:7687905"
/translation="MQQFTLLFLYLISAKYITGVDFSNLSLWNSNANYAFKPGP
YPTNVLGWSLDTGANPGDTFLNMPVCVRYKTTSTQSVLDADGVKATCOFYSGE
EFTFTSLTCTVNDALSKSIKAFCTVTLPIAFNVGGTGSFDLEDKSCFTAGTNTVTF
NDGDKDISIDVEPEKSTVDPISAYLXSRVMSLNKVTLLFVAPQCENGYTSGMTGFS
SNGDVAIDCSNIHIGITKGLNDNYPVSEFSYTKTCTSGNIOIKYONVPACRYPEI
DAISRTDQNYLAVTNDYTCAGRSQSKPFLRTWGYKNSDAGSNGIVIVATRVY
TDSSTAVTLPFPNSVDKTIIEILOPIPTITITISVGVYTSISKTAPIGETATVI
VDVYPTHTVTSEWGTGINTTITRNPTDSIDTVVQVPSNPPTVSTEYWSQSFAT
TTVTAPPGGTDSVIIIREPNHTVTTEYWSQSFATTTVTAPPGGTDSVIIIREPNP
TVTTEVWSQSFATTTVTAPPGGTDSVIIIREPNHTVTTEYWSQSFATTTVTAPP
GGTDSVIIIREPNHTVTTEYWSQSFATTTVTAPPGGTDSVIIIREPNHTVTTEY
WSQSFATTTVTAPPGGTDSVIIIREPNHTVTTEYWSQSFATTTVTAPPGGTDSVII
REPNHTVTTEYWSQSFATTTVTAPPGGTDSVIIIREPNHTVTTEYWSQSFATTT
VTAPPGGTDSVIIIREPNHTVTTEYWSQSFATTTVTAPPGGTDSVIIIREPNHTVT
TESNDITSIIPSRPHYVNTSDLSFESSMNTPTSISSDGLSLSTLTVTES
ETTESIGKSGVSVTETSVTITQTPNPLSSVTSLSLQSIIPSVSESKVTTSNGD
SSNVKSGVSVTETSVTITQTPNPLSSVTSLSLQSIIPSVSESKVTTSNGD
NOSGTHDSQSTSEIETVTSSTKVLPPVSSNTDLTSEPTNREOPTTLSTNSIT
EDLTQPTQDNGDNTSNPVTATSLASERENKSGSHESASTSLKPSMGNS
GLTSEIEATTSPEAPSPVSSGTDVTEPTDREOPTTLSTSKNSSEVATQ
ATWENKSGTDLTSLTGTASANSSELVTSGVGGVAVASNDOSHSTSVTN
SNSIVSNTPTQLSQVQVTSSTPTNFIASTYDGSISIIQHSWLXYLITLLSLFI"
1299..2370
/repeat_region
/rpt_unit=1299..1407
BASE COUNT 1170 a 899 c 592 g 1125 t
ORIGIN

Query Match 100.0%; Score 3786; DB 8; Length 3786;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTTCAACAATTACATGTTATCTCTATATTTGTCAATTGCAAGTGCACAAACATC 60
DB 1 ATGCTTCAACAATTACATGTTATCTCTATATTTGTCAATTGCAAGTGCACAAACATC 60
QY 61 ACTGGTGTGTTTGTAGTTTAAATTCATTAACTGGTCCCAATGCTGCTAAATATGCTTTC 120
DB 61 ACTGGTGTGTTTGTAGTTTAAATTCATTAACTGGTCCCAATGCTGCTAAATATGCTTTC 120
QY 121 AAAGGGCAGGATACCAACATGGAATGCTGTTTGGGTGGTCCCTAGATGGTACCAGT 180
DB 121 AAAGGGCAGGATACCAACATGGAATGCTGTTTGGGTGGTCCCTAGATGGTACCAGT 180
QY 181 GCCAATCCAGGGATACATTCACATTAATGATGCGATGTTGTTTAAATATGCTTTC 240
DB 181 GCCAATCCAGGGATACATTCACATTAATGATGCGATGTTGTTTAAATATGCTTTC 240
QY 241 CAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTTTCATTTTATTTCT 300
DB 241 CAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTTTCATTTTATTTCT 300
QY 301 GGTGAAGAATTCACAACATTTTCTACATTAACATGCTGTTGAACGAGCGCTTTGAAATCA 360
DB 301 GGTGAAGAATTCACAACATTTTCTACATTAACATGCTGTTGAACGAGCGCTTTGAAATCA 360

361 TCCATTAAAGGCATTTGGTACAGTTTACTTTACCAATTTGCAATTTGTTGGTGAACAGGT 420
361 TCCATTAAAGGCATTTGGTACAGTTTACTTTACCAATTTGCAATTTGTTGGTGAACAGGT 420
QY 421 TCATCAACTGATTTGGAGATTCTAAATGTTTACTGCTGTACCAATACAGTACACATTT 480
DB 421 TCATCAACTGATTTGGAGATTCTAAATGTTTACTGCTGTACCAATACAGTACACATTT 480
QY 481 AATGATGCTGATAAGATATCTCAATTTGAGTTTGAAGAGTCAACGTTTGTATCCA 540
DB 481 AATGATGCTGATAAGATATCTCAATTTGAGTTTGAAGAGTCAACGTTTGTATCCA 540
QY 541 AGTGCATATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTTCACAACTCTTTT 600
DB 541 AGTGCATATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTTCACAACTCTTTT 600
QY 601 GTGCACACCAATCTGAAATGGTTACACATCTCTGTAACATGGGGTCTCCAGTAGTAAC 660
DB 601 GTGCACACCAATCTGAAATGGTTACACATCTCTGTAACATGGGGTCTCCAGTAGTAAC 660
QY 661 GGTGACGTTGCTATTGATTTGCTCAAAATATTCATATTTGGTATCACAAAAGGATTAATGAT 720
DB 661 GGTGACGTTGCTATTGATTTGCTCAAAATATTCATATTTGGTATCACAAAAGGATTAATGAT 720
QY 721 TGGAAATATCCGGTTTCATCTGAATCATTTAGTTACATAAAGTCTGATCATCTTAATGGA 780
DB 721 TGGAAATATCCGGTTTCATCTGAATCATTTAGTTACATAAAGTCTGATCATCTTAATGGA 780
QY 781 ATTACAGATTAATATCAAAATGTACCTGCTGTTATCTGCTCATTTATTTGATGCTTATATT 840
DB 781 ATTACAGATTAATATCAAAATGTACCTGCTGTTATCTGCTCATTTATTTGATGCTTATATT 840
QY 841 TCTGCTACAGATTTAAACCAATATATCTTTAGCATATACCAATGATTTATGCTGCTGCG 900
DB 841 TCTGCTACAGATTTAAACCAATATATCTTTAGCATATACCAATGATTTATGCTGCTGCG 900
QY 901 AGTCTGCTGCAAACTAAACCTTTACATTTAGATGGAGTGGATACCAAAATAGTATGCC 960
DB 901 AGTCTGCTGCAAACTAAACCTTTACATTTAGATGGAGTGGATACCAAAATAGTATGCC 960
QY 961 GGATCTAACGGTATTTGCTCATTTGCTACAACTAGAACAGTTACAGACAGTACCACCTGCT 1020
DB 961 GGATCTAACGGTATTTGCTCATTTGCTACAACTAGAACAGTTACAGACAGTACCACCTGCT 1020
QY 1021 GTCATCTTACCATTCAATTCAGAGTTTGTATTAACCAACCAACCAACCAATTTGCAA 1080
DB 1021 GTCATCTTACCATTCAATTCAGAGTTTGTATTAACCAACCAACCAACCAATTTGCAA 1080
QY 1081 CCTATTCCAACTACCACTACCACTTCTATGTTGGTGGTACCTTCTCTATCTGACT 1140
DB 1081 CCTATTCCAACTACCACTACCACTTCTATGTTGGTGGTACCTTCTCTATCTGACT 1140
QY 1141 AAGACTGCACCAATTTGGTGAACAGTACTGTTTGTGATGTGCCATATCATCTACC 1200
DB 1141 AAGACTGCACCAATTTGGTGAACAGTACTGTTTGTGATGTGCCATATCATCTACC 1200
QY 1201 ACAACTGTTTACCAGTGAATGGACAGGAACAATCACTACCAACCAACCTCGTACCACCA 1260
DB 1201 ACAACTGTTTACCAGTGAATGGACAGGAACAATCACTACCAACCAACCTCGTACCACCA 1260
QY 1261 ACTGATCAATTTGACACAGTGGTGTACAAAGTCTCCAGTCCCAATCCAACTGTTAGTACT 1320
DB 1261 ACTGATCAATTTGACACAGTGGTGTACAAAGTCTCCAGTCCCAATCCAACTGTTAGTACT 1320
QY 1321 ACTGAATTTGGTCTCAGTCTCTTGTCTACCAACTACAGTACTGCTCCCTCCAGGTGCT 1380
DB 1321 ACTGAATTTGGTCTCAGTCTCTTGTCTACCAACTACAGTACTGCTCCCTCCAGGTGCT 1380
QY 1381 ACCGATCTGTTGATTTATCAGAGAGCCCAACCACTACTGTCCTACTTACTGAATATTGG 1440
DB 1381 ACCGATCTGTTGATTTATCAGAGAGCCCAACCACTACTGTCCTACTTACTGAATATTGG 1440
QY 1441 TCACAACTCTTGTCTACTACTACTACTGTTTACTGCTCTCCAGGTGCTGACTGACTAGTA 1500

|||||
Db 1441 TCACAACTCTTGTCTACTACTACTACTACTACTCTCTCCAGGTGGTACTGACTCAGTA 1500
QY 1501 ATTATCAGAGAACCACCAATCAACTGTCTACTACACCGAGTATGGTCTCAATCCTTT 1560
Db 1501 ATTATCAGAGAACCACCAATCAACTGTCTACTACACCGAGTATGGTCTCAATCCTTT 1560
QY 1561 GCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCACTAAATATATCAGAGAA 1620
Db 1561 GCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCACTAAATATCAGAGAA 1620
QY 1621 CCTCCAAACCACTGTCACCACTCAATATATGGTCCCAATCTTTACGCAACCACT 1680
Db 1621 CCTCCAAACCACTGTCACCACTCAATATATGGTCCCAATCTTTACGCAACCACT 1680
QY 1681 ACTGTGACTGCTCTCCAGGAGCACTGACTCACTAGTATATATCAGAGAACCCAAACCCAC 1740
Db 1681 ACTGTGACTGCTCTCCAGGAGCACTGACTCACTAGTATATATCAGAGAACCCAAACCCAC 1740
QY 1741 ACTGTGACTACTACTGAATACTGGTGCACAACTATGCCACCACTACCACCTGTAACCTGCA 1800
Db 1741 ACTGTGACTACTACTGAATACTGGTGCACAACTATGCCACCACTACCACCTGTAACCTGCA 1800
QY 1801 CCACGAGGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTGTCACCTACT 1860
Db 1801 CCACGAGGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTGTCACCTACT 1860
QY 1861 ACTGAGTATTTGGTCTCAATGTTGCTACTPACCACAACTGTAACCTGGTCCACCAAGTGGC 1920
Db 1861 ACTGAGTATTTGGTCTCAATGTTGCTACTPACCACAACTGTAACCTGGTCCACCAAGTGGC 1920
QY 1921 ACTGATACTGTTATCATTTAGGGAACCAACCAACCCAACTGTCAACCACTACTGTAATCTGG 1980
Db 1921 ACTGATACTGTTATCATTTAGGGAACCAACCAACCCAACTGTCAACCACTACTGTAATCTGG 1980
QY 1981 TCTCAATCATATGCAACCACTACTACCATTTACCGCTCCACCTGGTGAACCTGATACCGTT 2040
Db 1981 TCTCAATCATATGCAACCACTACTACCATTTACCGCTCCACCTGGTGAACCTGATACCGTT 2040
QY 2041 CTTTATCAGAGAGCCACCAACCACTACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
Db 2041 CTTTATCAGAGAGCCACCAACCACTACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
QY 2101 GCTACAAACCACTGTTACTGCACCACTGGTGAACCGATACCGTCTTATCAGAGAG 2160
Db 2101 GCTACAAACCACTGTTACTGCACCACTGGTGAACCGATACCGTCTTATCAGAGAG 2160
QY 2161 CCACCAACCACTACTGTCACCTACTGAATACTGGTCTCAATCATATGCTACAACCACT 2220
Db 2161 CCACCAACCACTACTGTCACCTACTGAATACTGGTCTCAATCATATGCTACAACCACT 2220
QY 2221 ACTGTTACTGCACCACTGCTACCGATACTGTTATCATTTAGAGAGCCACCAATTCCTCA 2280
Db 2221 ACTGTTACTGCACCACTGCTACCGATACTGTTATCATTTAGAGAGCCACCAATTCCTCA 2280
QY 2281 ACAGTTACTACTGTAATTTGGTGCACAACTATTTGCCACAACTGTCAGTTACTGCT 2340
Db 2281 ACAGTTACTACTGTAATTTGGTGCACAACTATTTGCCACAACTGTCAGTTACTGCT 2340
QY 2341 CCTCCAGGTGGTACTGACACTGTTATCTATGAAAGCATGTCAAGTTCAAGATTTCT 2400
Db 2341 CCTCCAGGTGGTACTGACACTGTTATCTATGAAAGCATGTCAAGTTCAAGATTTCT 2400
QY 2401 ACATCTCCCAATGATATACCACTGATCATTTCCCGTCTCTCATTTATGTCAC 2460
Db 2401 ACATCTCCCAATGATATACCACTGATCATTTCCCGTCTCTCATTTATGTCAC 2460
QY 2461 AGCACAACCTCCGATTTGTCACAACTTTGAATCTTCAATGCAATGATCTCTCTATC 2520
Db 2461 AGCACAACCTCCGATTTGTCACAACTTTGAATCTTCAATGCAATGATCTCTCTATC 2520
QY 2521 AGTAGTGATGGTATGTTGTTCTTCTACAACTTTGGTGTACTGAATCAGAGAACTAC 2580
|||||

Db 2521 AGTAGTGATGGTATGTTGTTCTTCTACAACTTTGTTACTGAATCAGAACTACATA 2580
QY 2581 GAACCTGATTTGCAGTGTGTAAGAGTGTCTAGATGTGTCAGTTCTTCTGCTGATTTGTC 2640
Db 2581 GAACCTGATTTGCAGTGTGTAAGAGTGTCTAGATGTGTCAGTTCTTCTGCTGATTTGTC 2640
QY 2641 ACAAATCCAGATAGCAATGAATCCTCAATGTAAGTACTGTTCTCTACTGCAAGTACA 2700
Db 2641 ACAAATCCAGATAGCAATGAATCCTCAATGTAAGTACTGTTCTCTACTGCAAGTACA 2700
QY 2701 ATGCTGATTTCACTTTCTTCAACTGATGGTATTTAGTGTCTACATCTTCTGATATGTTTCA 2760
Db 2701 ATGCTGATTTCACTTTCTTCAACTGATGGTATTTAGTGTCTACATCTTCTGATATGTTTCA 2760
QY 2761 AAATCAGGAGTATCAGTTACAACCGAAACTTCTGTTACAACCTATTCAAACTACTCCAAAC 2820
Db 2761 AAATCAGGAGTATCAGTTACAACCGAAACTTCTGTTACAACCTATTCAAACTACTCCAAAC 2820
QY 2821 CCATTTATCATCTTCTCAGTGACATCATTTGACTGACTGCTTCTTCAATTCCAAGTGTTCAGAA 2880
Db 2821 CCATTTATCATCTTCTCAGTGACATCATTTGACTGACTGCTTCTTCAATTCCAAGTGTTCAGAA 2880
QY 2881 AGTGAAGTAAAGTTACATTTACAAGCAATGGAGAGAACCAAGTGGTACTGATGATTTCA 2940
Db 2881 AGTGAAGTAAAGTTACATTTACAAGCAATGGAGAGAACCAAGTGGTACTGATGATTTCA 2940
QY 2941 CAATCTACTTCCACTGAAATTTGAAATTTGTAACCAACAGTGTCTACTAAAAGTTTACCACCT 3000
Db 2941 CAATCTACTTCCACTGAAATTTGAAATTTGTAACCAACAGTGTCTACTAAAAGTTTACCACCT 3000
QY 3001 GTCGTTTCTTCTAATCTGATTTGACTAGTGAACCAACCAATACCAGAGAACCAACCACT 3060
Db 3001 GTCGTTTCTTCTAATCTGATTTGACTAGTGAACCAACCAATACCAGAGAACCAACCACT 3060
QY 3061 ACATTTATCAACTACTTCAAACTCCATCCTGAAAGATATACCACATCTCAACCTACAGGT 3120
Db 3061 ACATTTATCAACTACTTCAAACTCCATCCTGAAAGATATACCACATCTCAACCTACAGGT 3120
QY 3121 GATAATGGAGACAATCTTTCATCAACCAATCCAGTTCCAACTGTGCAACCAAGTACTTTA 3180
Db 3121 GATAATGGAGACAATCTTTCATCAACCAATCCAGTTCCAACTGTGCAACCAAGTACTTTA 3180
QY 3181 GCATCTGCAAGTGAAGAGACAAACAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTG 3240
Db 3181 GCATCTGCAAGTGAAGAGACAAACAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTG 3240
QY 3241 AAACCAAGTATGGTGAATAATCTGGAATTAACCTACTTCTACTGAAATTTGAAGCTACAACA 3300
Db 3241 AAACCAAGTATGGTGAATAATCTGGAATTAACCTACTTCTACTGAAATTTGAAGCTACAACA 3300
QY 3301 ACCAGTCTACAGAAGCTCCATCAGCTGCTGTTTCTTCTGCTACTGATGTAACCTACTGAA 3360
Db 3301 ACCAGTCTACAGAAGCTCCATCAGCTGCTGTTTCTTCTGCTACTGATGTAACCTACTGAA 3360
QY 3361 CCAACTGATACTAGAGAACCACTACTACATTTATCAACTACTTCAAAACCAACCAAGTGA 3420
Db 3361 CCAACTGATACTAGAGAACCACTACTACATTTATCAACTACTTCAAAACCAACCAAGTGA 3420
QY 3421 CTGTTGCTACTACACAAGCTACTAATGAAATGGTGGTAAATCTCCATCAACTGATTTA 3480
Db 3421 CTGTTGCTACTACACAAGCTACTAATGAAATGGTGGTAAATCTCCATCAACTGATTTA 3480
QY 3481 ACATCAAGCTTTGACACAGGACCTCAGCATCTACAGTGTCTAATAGCAGAACTTGTACT 3540
Db 3481 ACATCAAGCTTTGACACAGGACCTCAGCATCTACAGTGTCTAATAGCAGAACTTGTACT 3540
QY 3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGTTCAAATGATCAATCAGATCTTACT 3600
Db 3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGTTCAAATGATCAATCAGATCTTACT 3600
QY 3601 TCTGTTACCACAGCAGCACTGATTTCTAATACCCCAACCAACTACATTTGAGTCAACAA 3660
Db 3601 TCTGTTACCACAGCAGCACTGATTTCTAATACCCCAACCAACTACATTTGAGTCAACAA 3660

```
Qy 3661 GTTACCTCATCTCACCCTTCAACCAACACATTCATTCGCTTCTACATACGATGGCTCTGGT 3720
|||||
Db 3661 GTTACCTCATCTCACCCTTCAACCAACACATTCATTCGCTTCTACATACGATGGCTCTGGT 3720
|||||
Qy 3721 TCTATTCCAAACATCTCTACTTGGTTGACGGTTTGATCACAATATTGCTCTGTTCAATT 3780
|||||
Db 3721 TCTATTCCAAACATCTCTACTTGGTTGACGGTTTGATCACAATATTGCTCTGTTCAATT 3780
|||||
Qy 3781 TAGTGA 3786
|||||
Db 3781 TAGTGA 3786

RESULT 2
CAU87956 3360 bp DNA linear PLN 02-JUL-1998
LOCUS Candida albicans agglutinin-like protein (ALS3) gene, complete cds.
DEFINITION U87956
ACCESSION U87956
VERSION U87956.1 GI:3273414
KEYWORDS
SOURCE
ORGANISM Candida albicans.
Candida albicans.
Eukaryota: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; microsporid Saccharomycetales; Candida.
REFERENCE
AUTHORS Hoyer,L.L., Payne,T.L., Bell,M., Myers,A.M. and Scherer,S.
TITLE Candida albicans ALS3 and insights into the nature of the ALS gene
family
JOURNAL Curr. Genet. 33 (6), 451-459 (1998)
MEDLINE 98309840
PUBMED 9644209
REFERENCE
AUTHORS Hoyer,L.L.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1997) Veterinary Pathobiology, University of
Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
source
1. 3360
/organism="Candida albicans"
/strain="1161"
/db.xref="taxon:5476"
/chromosome="R"
/map="5+0"
1. 3360
/gene="ALS3"
1. 3360
/gene="ALS3"
/codon_start=1
/trans_table=12
/product="agglutinin-like protein"
/protein_id="AAC39486.1"
/db.xref="GI:3273415"
/translations="MLQOYVLLIYLVSVAKTITGVNSFNSLTWSNATYVYKPGG
TPYNAVLSHSGTSAASPGDFTLAWPCVFKETTSQTSVDLTAHGKATYACQFQAGE
EFMTFSLTCTVNTLPSKALGTVLPLAFNVGGTSSVDLESKCFAGTNTVTF
NDGKQICINVDERSNVDPKGLTDSRVIPSLNKVSTLEVPQCCANGYTSGMTGFAN
TYGDVQIDSNVHIGITKGLDNWNPVSSSEFSYTKTSCSSNGIFITYKNVPAGYRPFV
DAYISATVNSYVLSYANEXYTCAGYQWAPFTLRWTCYRNSDAGSNGIVIVATRTVI
TDSATVTLUPDPNRDKTKTIELKPIPTTTTTSVGVTTSTYKTAPIGTATVI
VDIPIHTTNTVTSKWTGIPSTTHNPNTDSIDTVIVQVSPNPTVTTEYWSQSAT
TTITGPGNDVILIREPNHVTTEYWSQSYATTTTIFAPPGSDVSLIREPNP
TVITTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
GGDVLVIREPNHVTTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
SQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
REPNHVTTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
TITAPGETDVLIREPNHVTTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
TTEYWSQSYATTTTIFAPPGSDVSLIREPNHVTTEYWSQSYATTTTIVTAPP
ETKITSCGEGDKGSWSVSRIVTIPNNIETPMVNTVDSTESTESTSQSPSGIFS
ESGVSESTESTVTAQNPSTVPTTESEVFTTKGNENGPYESPTNKNKSSDENSE
FTTSTAESTDENETIATPGSVEASSPISSADETTVTVAESTSVIEQPTNN
GGKAPASATSSPSTTTANNDSVITGTSTNQSSQSOQYNSDQTQTTLSQOMTSSLVS
LHMLTFDGSQSVIQLHSTWLCGLITLLSLFI"
```

```
BASE COUNT 1029 a 804 c 520 g 1007 t
ORIGIN
Query Match 48.5%; Score 1834.4; DB 8; Length 3360;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 2393; Conservative 0; Mismatches 830; Indels 15; Gaps 5;
Qy 1 ATGCTTCAACAATTTACATTCGTTATTTCTATATTTGTCATTTGCAAGTGCAGAACAAATC 60
|||||
Db 1 ATGCTACAACAATATACATTCGTTACTCATATATTTGCGGTGGCACTGCAGAACAAATC 60
|||||
Qy 61 ACTGGTGTGTTTGTAGTGTGTTTAAATTCATTAACCTGGTCCAAATGCTGCTAAATGCTTTC 120
|||||
Db 61 ACTGGTGTGTTTCAACAGTGTGTTTAAATTCATTAACCTGGTCCAAATGCTGCTAAATGCTT 120
|||||
Qy 121 AAAGGGCAGATACCAACTTGAATGCTGTTTGGTGGTCTTGGTCTTGGTCTTGGTCTTGGT 180
|||||
Db 121 AAGGGACAGAAACCCCAACTTGAATGCTGTTTGGTGGTCTTGGTCTTGGTCTTGGTCTTGG 180
|||||
Qy 181 GCCAATCCAGGGATACATTCACATTCGAATATGCCATGTGTGTTTAAATATACACTTCA 240
|||||
Db 181 GCAAGTCCGGGAGATACATTCACATTCGAATATGCCATGTGTGTTTAAATATACACTTCT 240
|||||
Qy 241 CAACATCTGTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATTTCT 300
|||||
Db 241 CAACATCTGTGTTGATTTGACTGCTCATGCTGTTAAATATGCTACTTGTCAATTTTCAAG 300
|||||
Qy 301 GGTCAAGAATTCACAACTTTTCTTACATTAACATGTACTGTGAACGCGCTTTGAAATCA 360
|||||
Db 301 GGTCAAGAATTTAGACCTTTTCTTACATTAACATGTACTGTGAACGCGCTTTGACTCCA 360
|||||
Qy 361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTCATTAATGCTTGTGGTGGCAAGGT 420
|||||
Db 361 TCTATTAAAGCTTTGGGTACTGTCCACCTTACCACCTTGCATTCATTAATGCTGGAACGTGT 420
|||||
Qy 421 TCATCACTGATTTGGAGATTTCTAAATGTTTACTGCTGTGTACCAATACATGCTACATTT 480
|||||
Db 421 TCTTCTGTGTTGTTGGAAGATTTCTAAATGTTTACTGCTGTGTACCAATACATGCTACATTT 480
|||||
Qy 481 AATGATGCTGATAAGATATCTCAATTCATTCAGTTTGAAGTTTGAAGTTTGAAGTTTGAATCCA 540
|||||
Db 481 AATGATGGTGCAGAGAAATCTCTATTAAAGTTGATTTGAAAGTCAATATGTCGATCCA 540
|||||
Qy 541 AGTGCATATTGTTATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGGTGCACAACTCTTTT 600
|||||
Db 541 AAAGGTACTTAACTGATTCAGAGTTATACCAAGTCTCAACAAAGTGTCAACCTCTTTT 600
|||||
Qy 601 GTGCACCACAATCTGAAATGGTTACACATCTCGTACAAATGGGGTCTCCAGTAGTAAC 660
|||||
Db 601 GTTCACCAACAATGTGCAATGGTTACACATCTCGTCAATGGGATTCGCTTAACACTTAT 660
|||||
Qy 661 GGTGACGTTGCTATTGATTCGTTCAAAATATTCATATTGTTATCACAAAAGGATTAATGAT 720
|||||
Db 661 GGTGATGTTCAATGACTGTTCAAAATATTCATGTTGTTATACAAAAGGATTAATGAT 720
|||||
Qy 721 TGAATATATCGGTTTTCATCTGAATCATTTAGTTTACACTTAAACTTGTACATCTAATGGA 780
|||||
Db 721 TGAATATATCGGTTTTCATCTGAATCATTTAGTTTACACTTAAACTTGTCTAATGTT 780
|||||
Qy 781 ATTCAGATTAATATCAAAATGACCTGCTGGTTATCGTCCATTTATTTGATGCTTATATT 840
|||||
Db 781 ATCTTTATCAATATAAAATGTTCTCCCGGTTATCGTCCATTTGTTGACGCTTATATT 840
|||||
Qy 841 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATACCAATGATTAATGATTTGCTGGC 900
|||||
Db 841 TCTGCTACAGATGTTAAATTCGTACACCTTGTGCTGATGCTAATGATATACATTTGCTGGT 900
|||||
Qy 901 AGTCTGTCGAACTAAACCTTTTCACTTAAAGTGGACTGGATACAGAAATAGTAGTGCC 960
|||||
Db 901 GGTATTATGGCAACGTGACCTTTTCACTTAAAGTGGACTGGATACAGAAATAGTAGTGCT 960
|||||
Qy 961 GGATCTAACGGTATTGTCATTTGCTGCTACAACTAGACAGTTACAGACAGTACCACCTGCT 1020
|||||
```

Db	961	GGATCTAACGGTATGTTTATNTGTGGCTACTACCAAGACAGTTACAGACAGTACTACCGCT	1020
Qy	1021	GTCACTACTTTTACCATTCAATCCAAAGTGTGATATAAACCAAAACAATTCGAAATTTTGC	1080
Db	1021	GTGACCACCTTACCATTCCGATCCTTAACCCGCGACAAAACCTAAGACAATTGAATTTTGAA	1080
Qy	1081	CCTATTCCAACCACTACCATCAACAACCTTCATATGTTGGTGTGACTACTTCTCTATCTGACT	1140
Db	1081	CCTATTCCAACAACATACATCAACAACATCATATGTTGGTGTGACTACTTCTCTACCTGACC	1140
Qy	1141	AAGACTGCAACCAATTTGGTGAACACGCTACTGTTATTTGTTGATGTGCCATATCATACTAC	1200
Db	1141	AAAACCTGCACCAATTTGGGGAACCTGCTACTGTTATTTGTTGATATCCATATCACACTACC	1200
Qy	1201	ACAACTGTTTACCAGTGAATGGACAGAAACAATCACTACCAACCAACCTCGTACCAATCCA	1260
Db	1201	ACTACTGTTTACCAGTAAATTTGGACAGAAACAATTACTTCCACCAACACATACTAATCCA	1260
Qy	1261	ACTGATTCAATTTGACACAGTGGTGTGTACAAAGTTCCACTGCGCAAAATCCAAGTTTGTACT	1320
Db	1261	ACTGACTCAATAGACACTGCTCATTTGTACAAGTTCCACTGCGCAAAACCAACTGTTTACTACC	1320
Qy	1321	ACTGAATATTTGCTCTCAGTCCCTTCTCAACCACTACAGTTACTGCTCCTCCAGGTGGT	1380
Db	1321	ACTGAATATTTGGTCTCAATCATTTTTGCTACCAACCAACCATTACTTGACCACCAAGAAAC	1380
Qy	1381	ACCGACTACTGTGATTTATCAGACAGCCACAACCACTACTGTCATCTACTGTGAATATTTGG	1440
Db	1381	ACTGATCTACTGTTTAAATCAGAGAACCCGAAACCACTACTGTCATCAACCGAGTACTGG	1440
Qy	1441	TCACAATCCTTTTGCTACTACTACTACTGTTTACTGCTCCTCCAGGTGGTACTGACTCAGTA	1500
Db	1441	TCAGAATCTTATACACTACTACTAGTACTTCTACTGCTCCTCCAGGTGGAACCTGATTCACT	1500
Qy	1501	ATTATCAGAGAACCCAAATCCAACTGTCACTCAACCGAGTATTTGGTCTCAATCCTTT	1560
Db	1501	ATTATCAAGGAACCTCCAAATCCAACTGTCAACACTACCGAGTACTGGTCAGAACTTTAC	1560
Qy	1561	GCTACTACTACTACAGTTACTGCTCCTCCAGGTGTACTGACTCAGTAATTTATCAGAGAA	1620
Db	1561	ACTACCCTAGTACTTTCACCTCCCTCCAGGTGGAACCTGATTATTTATCAAGGAA	1620
Qy	1621	CCTCAAAACCAACTGTCAACCACTGAATTTGGTCCCAATCTTTACGCCAACCACTACT	1680
Db	1621	CCACCAAAACCACTGTAAACCAACACTGAATTTGGTCACAACTTTACACTTACCACCTACT	1680
Qy	1681	ACTGTCACTGCTCCTCCAGGAGCACTGACTCAGTAATTTATCAGAGAACCAACCAACAC	1740
Db	1681	ACTGTCAACGCTCCACAGGAGTACTGATAGTGTCTAGTCAGAGAACCAACCAACACAT	1740
Qy	1741	ACTGTCACTACTACTGAATACTTGGTCAACAATCATATGCCACCACTACCACTGTAACTGCA	1800
Db	1741	ACTGTTCAACTACCGAGTACTTGGTCACANCTTTACACTACACCACTGTTATTTGCC	1800
Qy	1801	CCACCAAGGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTGTCTACTACT	1860
Db	1801	CCACCAAGGTGGCACTGATTCGGTTATCATTTAGAGAAACCTCCAAATCCAACCTGTCAACA	1860
Qy	1861	ACTGAGTATTTGGTCTCAATCTGTTGGTACTACCACAAGTGTAACTGGTCCACCAAGTGGC	1920
Db	1861	ACTGAGTACTGGTCTCAATCTTACGCACTTACCACTACCATTAACGGTCTCCTCCAGGTGAG	1920
Qy	1921	ACTGATACTGTTATCATTTAGGGAACCAACCAACCACTGTCACCCTACTGATATCTG	1980
Db	1921	ACCGATACTGTCTTATTTAGAGAAACCAACCAACCACTGTAAACCAACCTGATATTTGG	1980
Qy	1981	TCTCAATCATATGCAACCACTACTTACCATTAACCGCTCCACCTGGTGAACCTGATACCGTT	2040
Db	1981	TCTCAATCTTATGCAACTACTACTACATCACTGCTCTCTCCAGGTGAACCGATACCGTT	2040
Qy	2041	CTTATCAGAGAGCCCAACCACTACTGTCTACTACTGATTAATCTGGTCTCAATCATAT	2100
Db	2041	CTTATTAGGAACCAACCAATCAGCTGTCACTACTACTGATTAATCTGGTCAACATCTTT	2100

Qy	2101	GCTACAACCACCACCTGTTACTGCACACACCTGGTGAACCGGATACCGTGTCTTATCAGAGAG	2160
Db	2101	GCTACAACCACAACCTGAACCTGACCCACCGGTGCTACTGACACTGTATTATCATTTAGAGAA	2160
Qy	2161	CCACAAACCATACTGCTACTACTAGTAATACTGGTCTCAATCATATGCTACAAACCACC	2220
Db	2161	CCACAAACCACACTGTCACCTACTACTGAGTATTGGTCTCAATCTTACGCAACCACCTACT	2220
Qy	2221	ACTGTTACTGCCACACCGGTGATACCGGATCTGTTATCATTTAGAGAGCCACCAAAATCCA	2280
Db	2221	ACCATTACCCTCCACTGGTGAGACCGATACCGGTTCTTATTAGGGAACCAACCAAAATCAC	2280
Qy	2281	ACAGTTACTACTAGTAATTTGGTCACAAATCATTTGCCAACAACCAACAGTTACTTGCT	2340
Db	2281	ACAGTTACTACTAGTAATTTGGTCACAAATCATATGCAACTACTACACCTATCATCGCA	2340
Qy	2341	CCTCAGGTGGTACTGCACACTGTGATTATCTATGA---AAGCATGTCAAAGTTCAAAGATT	2397
Db	2341	CCACCAGTGAACATGATACTGTTTAAATCAGAGAGCCACCAACCAACTGTTACCACC	2400
Qy	2398	TCTACATCTTCCAATGATATACCAAGTATCATTTCCATCATTTTCCGTCCTCATATTGTC	2457
Db	2401	ACCGAATACTGGTCTCAATCTATACCACTGCTACTACCGTTACTTGCAACCAAGGTGA	2460
Qy	2458	AACAGCAACACCTCCGATTGTCACATTTGAATCTTATCCATCATGAATAG---TCCCTACT	2514
Db	2461	ACTGATACTGTGATTTATTTATGACACCATGTCAGGTTCAGAAATTTCTTCAATTTCTCGT	2520
Qy	2515	TCATACAGTAGTGATGTATGTTGTTCTTCTACAACCTTTGGTTACTGAATCAGAACA	2574
Db	2521	CCTCATTTACCAACCATACAACCTTTGGTCTCAACCTTGGGTTATTGAAACAACAAACA	2580
Qy	2575	ACTACAGCACTTTGCAGTGATGTTGAAGAGTCTCTAGATGTGCCAGTCTTCTTGGT	2634
Db	2581	ATTACAGAACTACTGTGAAGGTGATAAGGTTGTTCTTGGGTTTCTGTTTCTTACTCGT	2640
Qy	2635	ATTGTCACAAATCCAGATAGCAATCCTCAATCGTAACCTAGTACTGTTTCTTACTGCA	2694
Db	2641	ATTGTCACAATTCCTAATAATATCGAAACTCCTATGGTTACTTAATACTGTTGATCTTACA	2700
Qy	2695	AGTCAATGTCGATTTCTTCTTCACTGATGTTATTAGTGTACATCTTCTTGATAAT	2754
Db	2701	ACCAGAACTCCACTTCACAACTCCCACTC---TGGTATTTTTCAGAGTCAGGAGTATCT	2757
Qy	2755	GTTTCAAAATCAGAGTATCAGTTACAAACCGAAACTCTGTTACAACCTATTTCAAACTACT	2814
Db	2758	GTTGAAACAGAACTCTTACTGTTACTACTGCTCAACAACAAATCCCAAGGTTTCCAACT	2817
Qy	2815	CCAAACCCATTCATCTTCAGTGACATCATTTGACTACTGAGTGTCTTCAATTTCCAAGTGT	2874
Db	2818	GAAGAGTGAGGTGTGTTTACTACTAAGGAAACACGAAATGGTCTTATGAATCACC	2877
Qy	2875	TCAGAAAGTGAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAGGTGACTCAT	2934
Db	2878	TCTACTAATGTGAATCAAGTATGGATGAAGAACTCTGAATTTA---CTTACTCCACAGCT	2934
Qy	2935	GATTACAATCTACTTCCACTGAAATGAA---ATTGTAAACCAACCGATTCTTACTAAAGTT	2991
Db	2935	GCTTCCACTTCTACTGATATTGAAATGAACCATAGCAACAACCGGTTCCGTGGAAGCT	2994
Qy	2992	TTACCACTGCTGTTTCTTAAATGATGTTGACTAGTGAACCAACAANATACCAGAA	3051
Db	2995	TCATCGGCTATCATTTCTTAGTGTGATGAACACTACTGTTTACTACTACTGCTGNA	3054
Qy	3052	CAACCAACTACATTTCACTTCTCAAACCTCCATCACTGAAGATATCACCACATCTCAA	3111
Db	3055	TCACCACTGCTATGAACAACCAACCATTAATATGGTGGTGAAGCCCATCTGCC	3114
Qy	3112	CCTACAGTGATAATGAGACAATACTTTCATCAACCAATCCAGTTTCCCACTGTGGCAACA	3171
Db	3115	ACTTCATCTCCACTACAACCTACAACCTGCTTAATGATGACTCTGTATTACTTGGTACAACA	3174

```

Qy 3172 AGTACTTACGATCTGCAAGTGAAGAGACACAAAGCGGTCTCTCATGAATCAGCAT 3229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3175 TCAACCAACCAATCTCAATCTCAATATAATATCTGTATACCAACCAACTACAT 3232

RESULT 3
AF051313 4383 bp DNA linear PLN 01-AUG-2000
LOCUS Candida albicans agglutinin-like cell surface protein (ALS8) gene,
DEFINITION complete cds.
ACCESSION AF051313
VERSION AF051313.2 GI:9625353
KEYWORDS
SOURCE Candida albicans.
ORGANISM Candida albicans.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 4383)
AUTHORS Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
TITLE Sequence of the hypha-specific, agglutinin-like cell surface
protein, ALS8 from Candida albicans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4383)
AUTHORS Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-1999) Molecular and Cell Biology, University of
Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen
AB25 2ZD, UK
COMMENT On Aug 1, 2000 this sequence version replaced gi:4105850.
FEATURES
source
    1..4383
     /organism="Candida albicans"
     /strain="ATCC10261"
     /db_xref="ATCC:10261"
     /db_xref="taxon:5476"
     566..>3840
     /gene="ALS8"
     566..>3840
     /gene="ALS8"
     /product="agglutinin-like cell surface protein"
     /note="minor form"
     574..579
     /gene="ALS8"
     /note="E-box"
     /bound_moiety="Efg1p"
     599..603
     /gene="ALS8"
     612..>3840
     /gene="ALS8"
     /product="agglutinin-like cell surface protein"
     /note="minor form"
     620..>3840
     /gene="ALS8"
     /product="agglutinin-like cell surface protein"
     /note="major form"
     630..>3840
     /gene="ALS8"
     /product="agglutinin-like cell surface protein"
     /note="minor form"
     697..3840
     /gene="ALS8"
     /function="Involved in cell-cell adhesion"
     /note="hypha-specific cell surface glycoprotein; similar
     to Candida albicans strain 1161 agglutinin-like protein
     encoded by ALS3 gene sequence presented in GenBank
     Accession Number U87956; contains two fewer internal
     repeated units than the Candida albicans strain 1161 ALS3
     gene"
     /codon_start=1
     /transl_table=12
     /product="agglutinin-like cell surface protein"
     /protein_id="AAD02580.1"
     /db_xref="GI:4105851"

```

```

/translation="MLQQYTLIIYLSVATAKITITGVFNSFLTWSNAATYHKGP
TPTNAVLGWSLDGTSASPDGFTLNPCVFKFTTSOTSDVLTAHKVKTACQOAGE
EMFTFSTLCTVSNTLTSPKALGTVTLPVAFNVGGTSSVDLEDKCTACGTNTVTF
NDGKKISINVDPERSNVDPKGYLDSRVIPSLNKVSTLFLVAPQCANGYSGTMGFAN
TYGDVDCSNHVGITKGLNDWNPVSEFSYTKCSNGIFITYNKVPAGYRFPV
DAYISADVNSYLSYAGNYCAGWQAPFTLRWTGRNSDAGSNGLIVATRTV
TOSTAVTLPFDPNKDTKITILKPIPTTTTISYGVGTTSTYKTKAPIGTATVI
VDIPIHTTSTKMTGTTTSTHTNPDSIDTVIQVPSNPVTVTTEYWSQSAT
TTTINCPGNTDVLIREPNNHTVTEYWSYTTSTFTAPPGCDSVILKEPNP
TVTTEYWSYTTSTFTAPPGCDSVILKEPNPVTVTTEYWSYTTTFTVTPAP
GGTDLVIREPNNHTVTEYWSYTTTIAAPPGDVTIIRPPNNHTVTTEY
SOSYATTTTAPPGEDTVLIREPNNHTVTEYWSYATTTTAPPGEDTVLI
REPNNHTVTTEYWSYATTTTAPPGDTVLIREPNNHTVTEYWSYATTT
TVTAPPGDVTIIRPPNNHTVTEYWSYATTTTAPPGDVTIIRPPNNHTV
GCSWVSRTIVTIPNNITPMVNTVDSTSTSTQSPGIFSEGSVETSESTV
TAQNPSPVPTTESEVEFTKGNNGPYEESPCHKSSMDENSEPTTSTAASSTDL
NATIATGSVEASSPIISSADETTITTAESTSVIEOPTNNGGKAPASTSPE
TTTANNDSVITGTTSTNQSSQSSTNQTTLSQOMTSSLSVLSHMLTTFDGGSGVI
QHSWLCGLITLLSLFI"
BASE COUNT 1359 a 922 c 705 g 1397 t
ORIGIN

Query Match 45.1%; Score 1708.6; DB 8; Length 4383;
Best Local Similarity 84.8%; Pred. No. 1.1e-298;
Matches 1918; Conservative 0; Mismatches 349; Indels 0; Gaps 0;

Qy 1 ATGCTTCAACAATTTACATTTGTTATTCCTATATTTGTCATTTGCAAGTGCAGAACACAATC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 ATGCTACAACAATATACATTTGTTACTATATATTTGTCGTTGCGACTGCGAAGACAATC 756

Qy 61 ACTGGTGTGTTTTGATAGTTTTAAATTCATTAACCTGGTCCAAATGCTGCTAATATATGCTTTC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 ACTGGTGTGTTTTCAACAGTTTTAAATTCATTCAGTTGGTCTAATGCTGCTACTTATCATAT 816

Qy 121 AAAGGCCAGGATACCAACTTGAATGCTGTTTGGTGGTGGCTTAGTGTGCTTACAGT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 AAGGGACAGGAACCCCAACTTGGATGCTGTTTGGTGGTGGCTTAGTGTGCTTACAGT 876

Qy 181 GCCAATCCAGGGATACATTCACATTCGAATATGCCATGCTGTGTTTAAATATACATCTCA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 GCAAGTCGGGGAGATACATTCACATTCGAATATGCCATGCTGTGTTTAAATTTACTACTTCT 936

Qy 241 CAAACATCTGTTGATTTAACTGGGATGGTGTAAATATGCTACTTGTCTCAATTTTATCT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 937 CAAACATCTGTTGATTTGACTGCTCATGGTGTAAATATGCTACTATGCTCAATTTTCAGGCA 996

Qy 301 GGTGAAGAATTCACAACTTTTCTACATTAACATGACTGTGAACGACGCTTTGAAATCA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 GGTGAAGAATTTATGACCTTTTCTACATTAACATGACTGTGAGCAATACTTTGAGTCCA 1056

Qy 361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTCATTCATGTTGGTGGAAACAGT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1057 TCTATTAAAGGCTTTGGGTACTGTCACTTACCCTTACCACTTCAATGATGAGTGAACGTG 1116

Qy 421 TCATCAACTGATTTGGAGATTTCTAAATGTTTACTGCTGGTACCATACAGTCACATTT 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1117 TCTTCTGTTGATTTGGAAGATTTCTAAATGTTTACTGCTGGTACTTAACACAGTTTACATTT 1176

Qy 481 AATGATGTTGATAAAGATATCTCAATGATGTTGAGTTTGAAAGTCAACCGTTGATCCA 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1177 AATGATGTTGGAAGAAATCTCTATTAAATGTTGATTTTGAAGGTCAATATGTCGATCCA 1236

Qy 541 AGTGCAATTTTGTATGTTCCAGAGTTTATGCCAAGTCTCAATAGGTCACAACTCTTTTT 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1237 AAAGGGTACTTAACTGATTCAGAGTTTATACCAAGTCTCAACAAAGGTCAACTCTTTTT 1296

Qy 601 GTGGCACCACAATGTGAAATGTTTACACATCTGGTACATGTTGGGTTCTCCAGTAGTAAC 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1297 GTTGCAACCAATGTGCAAAATGTTTACACATCTGGTACATGTTGGATTCGTTAAGTATAT 1356

Qy 661 GGTGACCTTCTGATTTGCTCAAAATATTCATATTCGTTATCACAAAAGGATTAATGAT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1357 GGTGATGTTCAAAATGACTGTTCAAAATATTCATATTCGTTATCACAAAAGGATTAATGAT 1416

```


QY 541 AGTGCATATTTGCTATGCTCCAGAGTTATGCCAAGTCTCAATAAGGTCACCAACTCTTTTT 600
DB 867 AGTGGTATTTGACTACTTCCAGATTTATCGGAGTCTCAATAAAATGGTACTCTTTAT 926
QY 601 GTGGACCACCAATGTGAATGGTTACACATCTGGTACATGGGTTCTCCAGTAGTAAC 660
DB 927 GTGGACCACCAATGTGAACGGTTACACATCTGGTACATGGGTTCTCCAGTAGTTAT 986
QY 661 GTGACGTTGTATTTGTTGCTCAATATTCATATTTGTTATCACAAAAGGATTAATGAT 720
DB 987 GGGATGTTGTTGTTGCTCAATATTCATATTTGTTATTCAAAAGGATTAATGAT 1046
QY 721 TGGAAATATCCGGTTTCACTGAATCAATTTAGTTACACTAAACCTGTACATCTAAATGGA 780
DB 1047 TGGAAATATCCGGTTTCACTGAATCAATTTAGTTACACTAAACCTGTCTCTTTGGT 1106
QY 781 ATTCAGATTAATATCAAAATGTACCTCTGTTATCGTCCATTTATTTGATGCTTATAT 840
DB 1107 ATCTCTACATATCAAAATGTCTCTCGGTTATCGTCCATTTATTTGACGCTTATAT 1166
QY 841 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATATACCAATGATTTATCTTGTGCTGC 900
DB 1167 TCTCCCTCAGATATAACCAAGTATCAATTTGCTATATAAAATGACTATCTTGTGTTGAT 1226
QY 901 AGTGTCTGCAAGTAACCTTTTACCTTTAAGATGGAGTGGATACAGAATAGTATGCC 960
DB 1227 GATTATTTGGCAGCATGCACTTTTCACTTTAAATGGAGTGGATATAAAGATAGTATGCC 1286
QY 961 GGATCTAACGGTATTTGCTATGTTGCTACAACTAGAACAGTTACAGACAGTACCACCTGCT 1020
DB 1287 GGATCTAACGGTATTTGCTATGTTGCTACAACTAGAACAGTTACAGACAGTACCACCTGCT 1346
QY 1021 GTCACACTTTTACCAATCAATCCAAAGTGTGATATAAAACCAAAACAAATCGAAATTTGCAA 1080
DB 1347 GTCACACTTTTACCAATCAATCCAAAGTGTGATATAAAACCAAAACAAATCGAAATTTGCAA 1406
QY 1081 CCTATTCAACCACTACCATCACAACTTCATATGTTGTTGTTGCTACTTCTCTACTGACT 1140
DB 1407 CCTATTCAACCACTACCATCACAACTTCATATGTTGTTGTTGCTACTTCTCTACTGACT 1466
QY 1141 AAGACTGCACCAATTTGGTGAACAGTACTCTGTTATTTGTTGTTGCTATATCTACTACC 1200
DB 1467 AAGACTGCACCAATTTGGTGAACAGTACTCTGTTATTTGTTGTTGCTATATCTACTACC 1526
QY 1201 ACAACTGTTACAGTGAATGACAGAGAACATCTACTACCACCAACCTCGTACCACCACTCA 1260
DB 1527 ACAACTGTTACAGTGAATGACAGAGAACATCTACTACCACCAACCTCGTACCACCACTCA 1586
QY 1261 ACTGATTCATTTGACACAGTGGTGTACAGTCCACTGCCAATCCCACTGTTAGTACT 1320
DB 1587 ACTGATTCATTTGACACAGTGGTGTACAGTCCACTGCCAATCCCACTGTTAGTACT 1646
QY 1321 ACTGAATATTTGGTCTCAGTCTCTTCTGCTACACCACTACAGTTACTGCTCCCTCCAGTGGT 1380
DB 1647 ACCAGTATTTGGTTCAGAGTCTATTTACTAGTACTACTACAACTACCAACTCTTTAAAGGT 1706
QY 1381 ACCGATCTGTGATATACAGAGGCCAACCAACCACTACTGCTACTACTGCTGAATATTTGG 1440
DB 1707 ACAGACTCAGTCAATTTGATAGAACACCAACCACTACTGTTTACCACACCGAGTTTGG 1766
QY 1441 TCACAACTCTTTGCTACTACTACTGTTTACTGTTCTCTCCAGTGGTACTGACTCAGTA 1500
DB 1767 TCAGAGTCAATTTGCCACTACTGAGACCAATCACAGCAAGAGGACAGAGTCCAGTC 1826
QY 1501 ATTATCAGAGAACCAACCAATCCAACTCTCACTACAACTGAGTATGTTGCTCAACTCCTTT 1560
DB 1827 ATTGTTAGAGAACCAACCACTCTACTGCTACTGCTACCACTGAGTTTGGTTCAGAGTCTGAT 1886
QY 1561 GCTACTACTACTAGTATTTACTGCTCTCCAGGTTGTTACTGCTAGTCAATTTATCAGAGAA 1620
DB 1887 GCTACTACTGAGACCACTACCAATGTTGCTCCAGAGAACAGACTCAGTCAATTTATGAGAA 1946
QY 1621 CCTCCAAACCACTGTCACCACTCACTGAATATTTGGTCCCAATCTTACCGCAACCACT 1680

DB 1947 CCACACACCCCTACTGTGACTACCAACCAAGTTTGGTCAGAGTGGTACTACTGAG 2006
QY 1681 ACTGTGACTGCTCTCCAGGAGGCACTGACTCAGTAATTTATCAGAGACCAACCAAC 1740
DB 2007 ACCATCACCAACAAGCCAGAGGACAGACTCAGTCATTGTTCAAGGAACCAACCCCT 2066
QY 1741 ACTGTCACTACTGTAATCTGTTCAATCATATGTCACCACTTACCACCTGTAACGCA 1800
DB 2067 ACTGTGACTACCAAGGATTTGGTCAGAGTCTGATGCTACTACTGAGACCACTCA 2126
QY 1801 CCACAGGTGTTGACTGACACTGTTTATCATTAGAGAGCCACCAACCACTGTTCACTACT 1860
DB 2127 GGTCCAGAAAGAACAGACTCAGTCAATGTTAGAGAACCAACCAACCCCTACTGCTACT 2186
QY 1861 ACTGATTTGCTCTCAATCTGTTGCTACTACCACTGTAACCTGTCACCAAGTGGC 1920
DB 2187 ACCGATTTGCTCAGAGTCTGATGCTACTGTAACCAATTTACTACAGGCCACTTGGT 2246
QY 1921 ACTGATCTGTTATCTAGGAAACCAACCACTGTCACCACTACTG 1972
DB 2247 ACTGATCTGTTATCTATCATGATCAATTTGGAAGAACTGCTTCTACTACTG 2298

RESULT 6
AF068866 3813 bp DNA linear PLN 29-JAN-2001
LOCUS
DEFINITION Candida albicans agglutinin-like protein (ALS5) gene, ALS5-1
allele, complete cds.
ACCESSION AF068866
VERSION AF068866.1 GI:4903268
KEYWORDS
SOURCE Candida albicans.
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS Hoyer, L.L. and Hecht, J.E.
TITLE The ALS5 gene of Candida albicans and analysis of the Als5p
N-terminal domain
JOURNAL Yeast 18 (1), 49-60 (2001)
MEDLINE 21064501
PUBMED 11124701
REFERENCE 2 (bases 1 to 3813)
AUTHORS Hoyer, L.L., Ho, M. and Hecht, J.E.
TITLE The ALS5, ALS6 and ALS7 genes of Candida albicans
JOURNAL Unpublished
AUTHORS Hoyer, L.L., Ho, M. and Hecht, J.E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) Veterinary Pathobiology, University of
Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL
61802, USA.

FEATURES
source

Location/Qualifiers
1..3813
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="3"
/note="Sfil fragment p"
1..3813
/gene="ALS5"
/allele="ALS5-1"
1..3813
/gene="ALS5"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAD32849.1"
/db_xref="GI:4903269"
/translation="MIQPTLLFLYLSFATAKAITGIENSIDSLTWSNAGNYAFKPGP
YPTWNAVGLNSLGTSTANPQFVFNKPKFTASOKSVDLTADGVKATPCQYSGE
EFTFTSLTCTVNDALKSSIKAFGTVTLPIAFNVGVTGSDLEDLSKCFTAGINTVTF

NDGSKLSIAVNFESTVDRSGYLTTSRFPMSLNKIATLYVAPOENGVTSGTMGFST
SYGDVAIDCSNVHIGISKGVNDNHPVTSESEYTKSCSFGISITUNVPAQYRPF1
YATSPSDNQLVSKYKNDYTCVDYWOHAPFTLKWGYKNSDAGSNGIYVATRTV
TDSITAVTLPNPSVDKTKIEILQPIPTTITTSYGVTSYKTAPEIGETALI
VDVPYHTTITVTSKWGTGITMTTTRTNPDSIDTVVQVPLNPPTTTTQWSESFIS
TTTTTHKEGSDSVIVKPHNPVTTFTEFWSYATTEITITPLGTDIVHDPLEE
SSSTAIEDSDNISSSAQESSVQSSIVGLSSSDIPLSMDPSSSSTGLTLEE
SSTVSYSDSSSELSTFSSSEYSSSIDTTFNFDSSSDLESTISITWSSSIDAQ
SSOSVOSVNSISNTOETSSSGESNTSVTDILVSDASSILNSDISYVPSSTISL
SDDPHTIAGBPDSSSSSIASSTVEISDLVSLTSDPTSSRDSSSSLSSSSPSFD
ESDLSASSFTLVPASFSSSSSLIYHYVNSITTHASESSSVASPSVASES
ANDHTLSESDTTSIIIGTDSSTVFCRNDGCVITGTTSSIDSEQSDVTTS
SFVASSSTAEQSTIDNPNDSSQTSASSTKSSVSDTVVNWILLSESTLSDD
GTSDTSSITNSDGTINAGSHKSTKESIOKTGVTLSSSYLSLSTLSSTDI
TIELTJLTITIEDNEPNTFTSPSSHEIFSSDNLVSKOVDRSTIKTSPYTDVTT
VSSLSVHSEASTATLGSPSNVASTPLNTATSLRSSSNHATESGIVKSEASV
EAPSPPTSDNRLPSSTEEAGITANGSTNNLITQESVAAQPSVSSHSSELST
STFDNSAAVDKPSKSIIESIMNPDSTNETNGFIATLSQAQVPSSSHSSELST
TKAKTDDASMGDSASNSQPTLLIQOVATSYNOPLITTVAGSSSATKHPSELKFI
SVALPFFL

BASE COUNT 1172 a 819 c 641 g 1181 t

ORIGIN

Query Match 31.4%; Score 1188.8; DB 8; Length 3813;
Best Local Similarity 78.8%; Pred. No. 7.9e-205;
Matches 1418; Conservative 0; Mismatches 382; Indels 0; Gaps 0;

QY 1 ATGCTTCAACAATTACATGTTATTCCTATATTCTGCAATTGCAAGTGCAGAGACAATC 60
DB 1 ATGATTCAACAATTACATGTTATTCCTATATTCTGCTGTTGCGACTGCAAGGCGATC 60

QY 61 ACTGGTGTGTTTGAATGTTTAACTTAATCTGCTTAACTGCTGCTAAATATGCTTTC 120
DB 61 ACTGGTGTGTTTGAATGTTTAACTTAATCTGCTTAACTGCTGCTAAATATGCTTTC 120

QY 121 AAAGGCCAGGATACCAACTTGGAAATGCTGTTTGGGTTGGTCTTATAGATGCTACAGT 180
DB 121 AAAGGCCAGGATACCAACTTGGAAATGCTGTTGGGTTGGTCTTATAGATGCTACAGT 180

QY 181 GCAATCCAGGGATACATTCATGTAATATGCCATGCTGTTTAAATATACACTTCA 240
DB 181 GCAATCCAGGGATACATTCATGTAATATGCCATGCTGTTTAAATATACACTTCA 240

QY 241 CAACATCTGTTGATTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300
DB 241 CAACATCTGTTGATTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300

QY 301 GGTGAAGATTCACAATCTTCTACATTAACATGACTGTGAACGACGCTTTGAAATCA 360
DB 301 GGTGAAGATTCACAATCTTCTACATTAACATGACTGTGAACGACGCTTTGAAATCA 360

QY 361 TCATTAAAGCATTTGCTGACGTTACTTTACCAATTTGCAATTCATGTTGGTGAACAGT 420
DB 361 TCATTAAAGCATTTGCTGACGTTACTTTACCAATTTGCAATTCATGTTGGTGAACAGT 420

QY 421 TCATCAACTGATTGGAAGATTCTAAATGTTTACTGCTGGTACCAATACAGTCACATTT 480
DB 421 TCATCAACTGATTGGAAGATTCTAAATGTTTACTGCTGGTACCAATACAGTCACATTT 480

QY 481 AATGATGGTGAATAAGATATCTCAATGATGTTGAGTTGTTGAAAGTCAACCGTTGATCCA 540
DB 481 AATGATGGGAGTAAAAAGCTCTCAATGCTGTTAAATTTGAAAGTCAACAGTTGATCCA 540

QY 541 AGTGCATATTGTTGCTTCCAGATTATGCCAAGTCTCAATAAGCTCACAACCTCTTTT 600
DB 541 AGTGGGTATTGACTATCTCCAGATTATGCGGAGTCTCAATAAAATTTGCTACTCTTAT 600

QY 601 GTGGCACCACAATGTGAAATGGTTACACATCTGGTACAATGGGTTCTCCAGTAGTAAC 660
DB 601 GTGGCACCACAATGTGAAACGGTTACACATCTGGTACAATGGGATTCTCCACTAGTTAT 660

QY 661 GTGACGTTGCTATTGATGCTCAAAATATTCATATTGGTATATCAAAAGGATTAATATGAT 720
DB 661 GTGACGTTGCTATTGATGCTCAAAATATTCATATTGGTATATCTCTGATATACCACA 1800

DB 661 GGGGATGTTGCTATTGACTGTTTCAAATGTACATATTGGTATTTCAAAAGGAGTAATGAT 720

QY 721 TGGAAATATCCGGTTCCTCATCTGAATCATTTAGTTACACTAAAACTTTGTACATCTAATGGA 780

DB 721 TGGAAATATCCGGTTCCTCATCTGAATCATTTAGTTACACTAAAACTTTGTACATCTAATGGA 780

QY 781 ATTCAAGTAAATATCAAAATFACCTGCTGGTGTATTCGTCCTATTTATTGATGCTTATATT 840

DB 781 ATCTCTATCATATCAAAATGTTCTCGCGGTATTCGTCCTATTTATTGACGCTTATATT 840

QY 841 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATATACCAATGATTTACTTGTGCTGGC 900

DB 841 TCTCCCTCAGATAAATAACCAAGTATCAATTCGTCGTAATAAATGACTATATCTTGTGTTGAT 900

QY 901 AGTCGCTCTCAAAAGTAAACCTTTCACTTTAAGATGACTGGATACAGAAATGATGATGCC 960

DB 901 GATTATTGGCAACATGACCTTCCTTCACTTTAATAATGACTGGATATAGAAATGATGATGCC 960

QY 961 GGATCTAAACGGTATTGTCATTTGCTTACAACCTAGAACAGTTACAGACAGTACCACTGCT 1020

DB 961 GGATCTAAACGGTATTGTCATTTGCTTACAACCTAGAACAGTTACAGACAGTACCACTGCT 1020

QY 1021 GTCACACTTTTACCATTCAATCCAAAGTGTGATAAACCACCAATCGAAATTTTGCAA 1080

DB 1021 GTCACACTTTTACCATTCAATCCAAAGTGTGATAAACCACCAATCGAAATTTTGCAA 1080

QY 1081 CCTATTCCAAACCACTACCACCACTTCATATGTTGGTGGTGAATCTCTCTCTATCTGACT 1140

DB 1081 CCTATTCCAAACCACTACCACCACTTCATATGTTGGTGGTGAATCTCTCTCTATCTGACT 1140

QY 1141 AAGACTGCAACCAATTGGTGAACAGCTACTGTTTATTTGTTGATGTGCAATATCATCTATACC 1200

DB 1141 AAGACTGCAACCAATTGGTGAACAGCTACTGTTTATTTGTTGATGTGCAATATCATCTATACC 1200

QY 1201 ACAACTGTTTACCAGTGAATGGAGCAGGAACAATCACTACCACCACTCTGTACCAATCCA 1260

DB 1201 ACAACTGTTTACCAGTGAATGGAGCAGGAACAATTCAGATGACTACAACCTGTACCAATCCA 1260

QY 1261 ACTGATTCAATTCACACAGTGGTGTACAAAGTTCACAGTCCCAATCCAACTGTTAGTACT 1320

DB 1261 ACTGATTCAATTCACACAGTGGTGTACAAAGTTCACAGTCCCAATCCAACTGTTAGTACT 1320

QY 1321 ACTGAATATGCTCAGTCTCAGTCTTGTGTACAAACCACTACAGTTACTGCTCTCTCCAGTGGT 1380

DB 1321 ACCCAGTTTTGGTCAGAGTCATTTACTAGTACTACTACAATCACAACCAAGCCAGAAGGC 1380

QY 1381 ACCGATCTGTTGATATACAGAGCCACCACCAACCACTACTGTCACTACTACTGATATTGG 1440

DB 1381 ACAGACTCAGTCAATTCAGAGGAACCAACCAATCCTACTGTTACCAACAGAGTATTGG 1440

QY 1441 TCACAATCCTTCTGCTACTACTACTGTTGCTCTCCTCCAGTGGTACTGACTCAGTA 1500

DB 1441 TCAGATCATATGCCACTACTGTAACCAATTTACTACAGGGCCACTTGGTACTGATGATATC 1500

QY 1501 ATTATCAGAGAACCACCAATCCAACTGTCACTACCAACCGAGTATGGTCTCAATCCCTTT 1560

DB 1501 GTTATACATCATCCATTCGGAAGAACGTCTCTCTACTACTGCTATTTAGTGTGAGTGATCT 1560

QY 1561 GCTACTACTACTACAGTACTGCTCTCCAGTGGTGGTACTGACTGATTAATTTATCAGAGAA 1620

DB 1561 AATATTTCAAGCTCAGCTCAAGAATCATCCAGCTGTTGGTGGAGCTCATCAAGTATATGTT 1620

QY 1621 CCTCCAAACCCCACTGTCACCACCTGAAATATTGGTCCCAATTTTACGCAACCAACACT 1680

DB 1621 GGATTTGATCAAGTTTCAGATATACCAATTAAGTTTCAGACATGCCATCATCGAGTCAACT 1680

QY 1681 ACTGTGACTGCTCTCTCCAGGAGGCACTGACTCAGTAATTTATCAGAGAACCCACCAACAC 1740

DB 1681 GGGTTAACTACTAGTGTAGTCTCTACTGCTCAAGTTATGATAGCGATTATCAAGTAGT 1740

QY 1741 ACTGTCACTACTACTGTAATGACTGGTCAACATCATATGCCACCACTACTGATCACTGCA 1800

DB 1741 AGTGAGTTATCTACATTTTCTAGTCTGAGAGCTACTGCTCAAGTATCTCTGATATACCACA 1800

RESULT 7
AR044076
LOCUS AR044076 1071 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5817466.
ACCESSION AR044076
VERSION *AR044076.1 GI:5965541
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1071)
AUTHORS Hoyer, L.L., Livi, G.P. and Shatzman, A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 5817466-A 1 06-OCT-1998;
FEATURES Location/Qualifiers
source 1..1071
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN
Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTGCTACAAACCACTACA 1359
Db 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTGCTACAAACCACTACA 60
Qy 1360 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
Db 61 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
Qy 1420 GTCACTACTACTGAATATTGGTCAACATCCTTCTGCTACTACTACTGTTTACTGCTCT 1479
Db 121 GTCACTACTACTGAATATTGGTCAACATCCTTCTGCTACTACTACTGTTTACTGCTCT 180
Qy 1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCAACCAATCCAACTGTCACACACC 1539
Db 181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCAACCAATCCAACTGTCACACACC 240
Qy 1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 1599
Db 241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 300
Qy 1600 GACTCAGTAATTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 1659
Db 301 GACTCAGTAATTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 360
Qy 1660 CAATCTTACGCAACCACTACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 1719
Db 361 CAATCTTACGCAACCACTACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 420
Qy 1720 ATCAGAGAACCAACCACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 1779
Db 421 ATCAGAGAACCAACCACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 480
Qy 1780 ACCACTACCACTGTAACCTGACCAACCACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 1839
Db 481 ACCACTACCACTGTAACCTGACCAACCACTGTCAGTCTCTCCAGAGGCACTGACTCAGTAATT 540
Qy 1840 CCAACCACTGTCAGTACTACTGAGTATTGGTCTCAATCGTTTGGTACTTACCACT 1899
Db 541 CCAACCACTGTCAGTACTACTGAGTATTGGTCTCAATCGTTTGGTACTTACCACT 600
Qy 1900 GTAAGTGGTCCACCAAGTGGCACTGATCTGTTATCATTAGGGAACCAACCACT 1959
Db 601 GTAAGTGGTCCACCAAGTGGCACTGATCTGTTATCATTAGGGAACCAACCACT 660
Qy 1960 GTCACCACTACTGTAATCTGCTCAATCATATGCAACCACTACTACCACTTACCGTCCA 2019
Db 661 GTCACCACTACTGTAATCTGCTCAATCATATGCAACCACTACTACCACTTACCGTCCA 720

Qy 2020 CCTGGTGAAGTATACCGTTCTTTATCAGAGAGCCACCAACCACTACTGTGCTACTACT 2079
Db 721 CCTGGTGAAGTATACCGTTCTTCATCAGAGAGCCACCAACCACTACTGTGCTACTACT 780
Qy 2080 GAATACTGGTCTCAATCATATGCTACACCACTACTGTTACTGCAACCACTGGTGAAC 2139
Db 781 GAATACTGGTCTCAATCATATGCTACACCACTACTGTTACTGCAACCACTGGTGAAC 840
Qy 2140 GATACCGTCTCTTATCAGAGAGCCACCAACCACTACTGTGCTACTACTGTAATCTGGTCT 2199
Db 841 GATACCGTCTCTTATCAGAGAGCCACCAACCACTACTGTGCTACTACTGTAATCTGGTCT 900
Qy 2200 CAATCATATGCTACACCACTACTGTTACTGCAACCACTGGTGGTACCGATGTTATC 2259
Db 901 CAATCATATGCTACACCACTACTGTTACTGCAACCACTGGTGGTACCGATGTTATC 960
Qy 2260 ATTAGAGAGCCACCAACCACTACTGTTACTGCAACCACTGTAATCTGGTGGTGGTCC 2319
Db 961 ATTAGAGAGCCACCAACCACTACTGTTACTGCAACCACTGTAATCTGGTGGTGGTCC 1020
Qy 2320 ACAACCACTACTGTTACTGCTCTCCAGGTGGTACTGACACTGTTGATTATC 2370
Db 1021 ACAACCACTACTGTTACTGCTCTCCAGGTGGTACTGACACTGTTGATTATC 1071

RESULT 8
I65632
LOCUS I65632 1071 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 5668263.
ACCESSION I65632
VERSION I65632.1 GI:2482202
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1071)
AUTHORS Hoyer, L.L., Livi, G.P. and Shatzman, A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 5668263-A 1 16-SEP-1997;
FEATURES Location/Qualifiers
source 1..1071
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN
Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGTGCTACAAACCACTACA 1359
Db 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGTGCTACAAACCACTACA 60
Qy 1360 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
Db 61 GTTACTGCTCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
Qy 1420 GTCACTACTACTGAATATTGGTCAACATCCTTCTGCTACTACTACTGTTTACTGCTCT 1479
Db 121 GTCACTACTACTGAATATTGGTCAACATCCTTCTGCTACTACTACTGTTTACTGCTCT 180
Qy 1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCAACCAATCCAACTGTCACACACC 1539
Db 181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCAACCAATCCAACTGTCACACACC 240
Qy 1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 1599
Db 241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCTCCAGGTGGTACT 300
Qy 1600 GACTCAGTAATTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 1659
Db 301 GACTCAGTAATTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 360

QY	1660	CAATCTTACGCAACCACTACTGTGACTGCTCTCCAGGAGGCACTGACTCAGTAATT	1719
Db	361	CAATCTTACGCAACCACTACTGTGACTGCTCTCCAGGAGGCACTGACTCAGTAATT	420
QY	1720	ATCAGAGAACCAACCACTGTCCTACTACTGTAATGCTGTCACATCATATGCC	1779
Db	421	ATCAGAGAACCAACCACTGTCCTACTACTGTAATGCTGTCACATCATATGCC	480
QY	1780	ACCACCTACCACTGTAATGCTGACCAACAGGCTGCTGACACTGTTATCATAGAGGCCA	1839
Db	481	ACCACCTACCACTGTAATGCTGACCAACAGGCTGCTGACACTGTTATCATAGAGGCCA	540
QY	1840	CCAAACCACTGTCCTACTACTGAGTATGCTCAATGCTGTTGCTACTACCAAACT	1899
Db	541	CCAAACCACTGTCCTACTACTGAGTATGCTCAATGCTGTTGCTACTACCAAACT	600
QY	1900	GTAACTGGTCCCAACAGTGGCACTGATAGTGTATCATATAGGAAACCAACCACT	1959
Db	601	GTAACTGGTCCCAACAGTGGCACTGATAGTGTATCATATAGGAAACCAACCACT	660
QY	1960	GTCAACCACTACTGAATGCTGCTCAATCATATGCAACCACTACTACCACTACCGCTCCA	2019
Db	661	GTCAACCACTACTGAATGCTGCTCAATCATATGCAACCACTACTACCACTACCGCTCCA	720
QY	2020	CCTGGTGAACCTGATACCGTCTTATCATAGAGAGCCACCAACCACTACTGTCCTACTACT	2079
Db	721	CCTGGTGAACCTGATACCGTCTTATCATAGAGAGCCACCAACCACTACTGTCCTACTACT	780
QY	2080	GAATAGTGGTCTCAATCATATGCTACACCACTGTTACTGCACCACTGCTGAAACC	2139
Db	781	GAATAGTGGTCTCAATCATATGCTACACCACTGTTACTGCACCACTGCTGAAACC	840
QY	2140	GATACCGTCTTATCATAGAGAGCCACCAACCACTACTGTCCTACTACTGAATGCTGCT	2199
Db	841	GATACCGTCTTATCATAGAGAGCCACCAACCACTACTGTCCTACTACTGAATGCTGCT	900
QY	2200	CAATCATATGCTACACCACTGTTACTGACCACTGCTGTCCTACTACTGTAATGCTGCT	2259
Db	901	CAATCATATGCTACACCACTGTTACTGACCACTGCTGTCCTACTACTGTAATGCTGCT	960
QY	2260	ATTAGAGAGCCCAATCAACAGTACTACTACTGTAATGCTGTAATGCTGTAATGCTGCT	2319
Db	961	ATTAGAGAGCCCAATCAACAGTACTACTACTGTAATGCTGTAATGCTGTAATGCTGCT	1020
QY	2320	ACAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2370
Db	1021	ACAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1071
RESULT 9			
CANALS2S1			
LOCUS		1404 bp DNA linear	PLN 16-OCT-1998
DEFINITION		Candida albicans agglutinin-like protein (ALS2) gene, 5' partial	
ACCESSION	AF024582		
VERSION	AF024582.1	GI:3598672	
KEYWORDS			
SEGMENT			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

JOURNAL	Submitted (11-SEP-1997) Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES	
Source	Location/Qualifiers
	1..1404
	/organism="Candida albicans"
	/strain="1161"
	/db_xref="taxon:5476"
	/chromosome="6"
	/map="6C"
gene	1..>1404
	/gene="ALS2"
	/note="agglutinin-like sequence"
	/allele="ALS2-2"
CDS	1..>1404
	/gene="ALS2"
	/codon_start=1
	/transl_table=12
	/product="agglutinin-like protein"
	/protein_id="AAC64237.1"
	/db_xref="GI:3598675"
	/translation="MLLQFLLLSLCVSVATAKVIITGVFNSFDSLWTRAGNVAYKGN RPTNAVLGWSLDGTSANPDFTLNMPCVFKFITDSTVLDLTAEVKVATQCFYSGE EFTTSSLLACTVSNLTLSIKALGTVTLPISENVGGTSSVDLESQCFKAGTNTVTF NGDKKISIDVDFEKTNEBDSGFFIASRLIPSINKVSIYVAPQCCANGTSGAMGFIV LGDDTTIDCSNVHVGITKGLNDWNPVSSDSUSYNKTSIGISITITVNPVAGYRFFV DYTYSVGNQRLRYNDYACVSSLSQKPFNLRLGRYNNSEANSNFVIVATRTVT DSTAVTTLFPNPSVDKTIIELOPITTTITTSYGVTTSTYKTAIGETATVIV DVPYHTTITVSEWGTGTITATTTTNTPTDSIDTVVQVPSNPVTSTTEVWSQSYATT TTVTPPGGCTDSVLIIEP"
repeat_region	1397..1404
	/note="first of multiple copies of 108bp imperfect repeat"
BASE COUNT	421 a 266 c 255 g 462 t
ORIGIN	
	Query Match 26.9%; Score 1016.6; DB 8; Length 1404;
	Best Local Similarity 83.2%; Pred. No. 1.2e-173;
	Matches 1170; Conservative 0; Mismatches 234; Indels 3; Gaps 1;
QY	1 ATGCTTCAACAATTTACATTTGTTATTCCTATATTTGTCATTTGCAAGTGGCAAGACAACT 60
Db	1 ATGCTTTTACAATTTTGTGCTAAGCCTCTGCTATCAGTGTCTACTGCAAAAGTTATT 60
QY	61 ACTGGTGTGTTGATGTTTAAATTAATTAAGTGGTCCCAATGCTGCTAATATGCTTTTC 120
Db	61 ACGGGTGTTTCAATAGTGTGTTGATTCGTTGACATGGAAGAGCTGTTAATGCTTAT 120
QY	121 AAAGGCCAGGATACCAACTTGAATGCTGTTTGGTGGCTTTAGATGGTACCACT 180
Db	121 AAGGCCCAAAATAGCAACTTGAATGCTGTTTGGGCTGCTTTAGATGGTACCTAGT 180
QY	181 GCCAATCCAGGGATACATTCACATTTGAATATGCTGTTTAAATATATCTACTTCA 240
Db	181 GCAATCCAGGAGACATTCACATTTCAATATGCTGTTTAAATATATCTACTTCA 240
QY	241 CAACATCTGTTGATTTAACTGCGGATGGTGTAAATATGCTGCTGCTGCTGCTGCTGCT 300
Db	241 CAACATCTGTTGATTTAACTGCGGATGGTGTAAATATGCTGCTGCTGCTGCTGCTGCT 300
QY	301 GGTGAAGAATTCACAACTTTTCTACATTAACATTAACATTAACATTAACATTAACATTA 360
Db	301 GGTGAAGAATTTACAACTTTTCTCAATTAATTAATTAATTAATTAATTAATTAATTA 360
QY	361 TCAATTAAGGCTTGGTACAGTTTACCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db	361 TCAATTAAGGCTTGGTACAGTTTACCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY	421 TCATCAACTGATTTGGAAGATTTTAAATGTTTACTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db	421 TCATCGGTGATTTGGAAGATTTTCAATGTTTAAAGGCTGCGCACCAACACAGTTTAC 480
QY	481 AATGATGTTGATAAAGATATCTCAATTTGATGTTTGAAGTTTGAAGTTTGAAGTTTGA 540

```
Db 481 AATGATGGTGATAAAAAATTTTCAATTGACGCTTGATTTTGAAAAAACAAGAGATGCA 540
Qy 541 AGTCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTT 600
Db 541 AGTGAATTTTCATAGCGTCAAGACTTATCCAGTATTACAAAGTTTCAATCAGTTAT 600
Qy 601 GTGCACACAAATGTGAATGGTTACACATCTGGTCAATAGGGGTTCTCCAGTAGAAC 660
Db 601 GTGCACACAAATGTGAATGGTTCACATCTGGTCAATAGGGGTTCTAGTTCTCACT 660
Qy 661 GGTGACGTTGCTATGATGCTCAATATTCATATTTGGTATCACAAAGGATTAATGAT 720
Db 661 GGTGACACTATGATGCTCAATATTCATATTTGGTATCACAAAGGATTAATGAT 720
Qy 721 TGAATATCCGGTTTCATCTGAATCAATTTAGTTACACTAAACTCTGTACATCAATGGA 780
Db 721 TGAATTTCCGGTATGCTGATTCATTAAGTTACAAATAAACTTGTTCATCAGGT 780
Qy 781 ATTGAGATTAATATCAAAATGATACCTGCTGGTTATCGTCCATTTATGATGCTTAT 840
Db 781 ATTTCTATACATATGAATGTCCTGGTATCGTCCATTTTACGATATATACT 840
Qy 841 TCTGCTACAGATGTTAACCAATATCTTAGCATATACCAATGATTACTTGTCTGGC 900
Db 841 CTGGTGTCAGGCCAGAACAGACA---ATTAAGATATATACTAATGATTGCGCTGTTGGT 897
Qy 901 AGTCCTCTGCAAAAGTAAACCTTTCACTTTTAAGATGGACTGGATACAGAAATGATGCC 960
Db 901 AGTCCTCTGCAAAAGTAAACCTTTCACTTTTAAGATGGACTGGATACAGAAATGATGCC 960
Qy 957 AGTTCCCTTACAAAGTAAAGCGCTTCAATTTAAGATTGAGAGGATACAAATAGTGAAGCT 957
Qy 961 GGATCTAACGGTATGCTATGTTGCTTCAACTGCTGGTTATCGTCCATTTATGATGCT 1020
Db 958 AATCTCAACGGTTTGTCAATGTTGCTTACAAACCGCAAGTACTGACAGTACTGCT 1017
Qy 1021 GTCACTACTTTTACCAATCAATCAAGTGTGATATAAACCACAAACAAATCGAAATTTGCAA 1080
Db 1018 GTCACTACTTTTACCTTTTATCCAAAGTTGTGACAAACCAACAAATCGAAATTTGCAA 1077
Qy 1081 CCTATTCACCACTACATACAACTTCATATGTTGGTGTGACTTCTCTATCTGACT 1140
Db 1078 CCTATTCACCACTACATACAACTTCATATGTTGGTGTGACTTCTCTATCTGACT 1137
Qy 1141 AAGACTGACCAATTTGTTGAACAGTACTCTTATTTGATGTGCCATATCATCTACC 1200
Db 1138 AAGACTGACCAATTTGTTGAACAGTACTCTTATTTGATGTGCCATATCATCTACC 1197
Qy 1201 ACAACTGTTACAGTGAATGGACAGGAACATCACTACCAACCAACTCGTACCAATCCA 1260
Db 1198 ACAACTGTTACAGTGAATGGACAGGAACATCACTACCAACTCGTACCAATCCA 1257
Qy 1261 ACTGATTCATTTGACAGTGGTGGTACAAGTTCCTACTGCCAAATCCAACTGTTAGTACT 1320
Db 1258 ACTGATTCATAGATGCTGCTGTTGCAAGTTCCTACTGCCAAATCCAACTGTTAGTACT 1317
Qy 1321 ACTGAATATGGTCTCAGTCTTGTCTACACCACTACAGTTACTGCTCTCCAGGTGGT 1380
Db 1318 ACTGAATATGGTCTCAGTCTATGCTACACCACTACAGTTACTGCTCTCCAGGTGGT 1377
Qy 1381 ACCGATCTGTGATTTATCAGAGGCCA 1407
Db 1378 ACTGACTCAGTCATATCAGAGGCCA 1404

RESULT 10
CALALS1
LOCUS
DEFINITION
  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial cds.
ACCESSION
  AF024580
VERSION
  AF024580.1
KEYWORDS
  1 of 2
SEGMENT
  Candida albicans.
SOURCE
```

```
ORGANISM
  Candida albicans
  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
  1 (bases 1 to 1404)
AUTHORS
  Hoyer, L.L., Payne, T.L. and Hecht, J.E.
TITLE
  Identification of Candida albicans ALS2 and ALS4 and localization
  of als proteins to the fungal cell surface
JOURNAL
  J. Bacteriol. 180 (20), 5334-5343 (1998)
MEDLINE
  98440424
PUBMED
  9765564
REFERENCE
  2 (bases 1 to 1404)
AUTHORS
  Hoyer, L.L.
TITLE
  Direct Submission
JOURNAL
  Submitted (11-SEP-1997) Veterinary Pathobiology, University of
  Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
  Location/Qualifiers
  1..1404
     /organism="Candida albicans"
     /strain="1161"
     /db_xref="taxon:5476"
     /chromosome="6"
     /map="6; C"
     /gene="ALS2"
     /codon_start=1
     /transl_table=12
     /product="agglutinin-like protein"
     /protein_id="AAC64235.1"
     /db_xref="GI:3598670"
     /translation="MLQFLLSLCVSATAKVITGVNFSFSLTWTRAGNYAKGN
     RPTWNAVGLMSLDGTSANPGDTITLNNPCVKFIDTQTSVDLTAEGVKATCFYSGE
     EFTFSSLKTSVNTLTSSIKALGTVLPISFNVGGTSSVDLESQCFKAGTNTVTF
     NQDGKKISIDVDEKTNEDASGYFIASRLIPSIKVSITYVAPQANGYTGSGMGFIV
     LCGDTTIDCSNVHVGITKGLNDWNPVSSDSLSYKNTCSSTGTSITVENVPAGYRPF
     DVTYSVGNQRLRYNDYACVGSLSQSKPFLNRLRGYNNSEANSNGFVAVATRTVT
     DSTAVTLPPEVSDTKTIEILOPPTTITTSYGVGTTSTKTPAGTGTATVIV
     DVPHTTTTISEWTGTTTTTTTTRNPTSDISIVVQVPSFNTVTITTEWSQSIAT
     TTVTAPPGGTDVSIIREP"
     repeat_region
     1297..1404
     /note="first of multiple copies of imperfect 108 bp
     repeat"
BASE COUNT
  422 a 268 c 253 g 461 t
ORIGIN
  Query Match 26.6%; Score 1006; DB 8; Length 1404;
  Best Local Similarity 82.7%; Pred. No. 9,8e-172;
  Matches 1163; Conservative 0; Mismatches 240; Indels 3; Gaps 1;
  Qy 1 ATGCTTCAACAATTTACATGTTTATCTCTATATTTGCAATTTGCAAGTGCACAAACAATC 60
  Db 1 ATGCTTTTACAATTTTGTGCTAAGCCTCTGTGTATCAGTTGCTACTGCAAAAAGTTATT 60
  Qy 61 ACTGCTGTTTGTAGTGTATTTAATTCATTAATCTGGTCCCAATGCTGCTAATATGCTTTC 120
  Db 61 ACGGCTGTTTTCATAGTGTGTTGATTCGTTGACATGGACGAGAGCTGTTAATATGCTTAT 120
  Qy 121 AAAGGGCCAGGATACCCAACTTGGAAATGCTGTTTGGTGGTCTGTAGATGGTACCAAGT 180
  Db 121 AAAGGCCCAATAGACCACTTGGAAATGCTGTTTGGCTGCTTTAGATGGTACTAGT 180
  Qy 181 GCCAATCCAGGGGATACATTCACATTTGAATATGCCATGTGTGTTTAAATATACACTTCA 240
  Db 181 GCAATCCAGGAGACACATTCACATTTGAATATGCCATGTGTGTTTAAATTTATTACCAT 240
  Qy 241 CAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCT 300
  Db 241 CAACATCTGTTGATTTGACGTCGAGGGTGTAAATATGCTACTGTCAGTTTATTTTCA 300
  Qy 301 GGTGAAGAATTCACAACCTTTTCTACATTAACATGCTACTGTGAACGAGCGCTTTGAAATCA 360
  Db 301 GGTGAAGAATTTACAACATTTTCTTCAATTAATGCTACTGTGACCAATCTATTACATCA 360
```


[illegible]

Db	26	TTTTCTCTCATTCGACAGATAGCAATGGCGAAAACATATATCGGGAGTTTTTTCACGAGTTTCA	85
Qy	83	ATTCAATTAACTTGGTCCCAATGCTGCTAAATTATGCTTTCAAAGGGCCAGGATACCCAACTT	142
Db	86	ACTCAATTGACCTATACTAATACTAGTTAACTACCCATATGGGGTCTTGGTTATCCAACT	145
Qy	143	GGATGCTGTTTTGGCTTGGTCTTAGATGGTTACAGATGCCAATCCAGGGGATACATCA	202
Db	146	GGACTGCTGTTTTAGTTGGAGCTTGGACGGAACACTAGCTAGTCCAGGTGATACATTTA	205
Qy	203	CATTGAATATGCCATGCTGTTTTAAATATACTTCTACAAACATCTGTTGATTAACTG	262
Db	206	CATTGGTCATGCCCTCGTTTTCAATTTATTACCACACAACCTCAGTAGACTTAACG	265
Qy	263	CCGATGGTGTAAATATGCTACTTGTCAATTTTATTCTGGTGAAGAATTACAACTTTTTT	322
Db	266	CTAATGGTGTCAAGTAGTACCAATGACTTTTCCATGTCAGGGGAAGACTTTACTACTTTTT	325
Qy	323	CTACATTTAAACATGACTGTGAACGAGCGTTTGAATCATCTCAATTAAAGSCATTTGGTACAG	382
Db	326	CAAGTAGTGTGTTAGTAAATAATGGGCTATCTTCAATATACAGAGCGTTTGGTACCG	385
Qy	383	TTACTTTACCAATTTGCATTCATGTTGGTGGAAAGGTTTCATCACTGATTTTGAAGATT	442
Db	386	TCAGGCTACCAATTTCAATTCATGTTGGTGGAACTGGTTTCATCTGCTCAACATTCAGATT	445
Qy	443	CTAAATGTTTTTACTGCTGGTACCAATACAGTCCACATTTAATGATGGTGTGAATAAGATATCT	502
Db	446	CAAAAGTGTTCACGTGCTGGAAAGCACTGTAAACATTTTACAGACGGCGATCCAAAAATTT	505
Qy	503	CAATTTGATGTGGATTGAAAAGTCAACCGTTTGATCCAAAGTGCATATTATTGTATGCTTCCA	562
Db	506	CTACTACAGTCAATTTTCCCTTAAGACTCCACAATCATCTAGTAGCTTGGTTTTATTTCGCAA	565
Qy	563	GAGTTATGCCAAGCTCTCAATAAGGTCACAACACTTTTTTGTGGCACCACAATGTCAAAATG	622
Db	566	GGGTATTTCCAAGCTGTGATAAATTATCTAGTCTTGTGTTGCTTCTCAGTGTACTGCTG	625
Qy	623	GTTACACATCTGTTACAATGGGGTTCTCCAGTAGTAACGGTGCAGTGTCTATTGATTGCT	682
Db	626	GATATGCTCCGGTGTCTCGGATTTTCAGCAACAAAAGATGATGTGACAAATTGATTGTT	685
Qy	-683	CAAAATTATCATTTGTTATCACAAAAGGATTTAAATGATTTGGAATATATCCGGTTTCATCTG	742
Db	686	CTACTATACATTTGGGAATTAACAAATGGTTTGAATAGTTTGGAAATATGCCAGTATCATCAG	745
Qy	743	AATCATTTAGTTTACACTAAACCTTGTCATCTCTAATGGAAATTCAGATTAAATATCAAAATG	802
Db	746	AATCATTTTCTTACACAAAACCTTGTCACACCAACAGTTTTATTATTACTTTATGAAAATG	805
Qy	803	TACCTGCTGGTTATGTCGCAATTTATTGATGCTT-----ATATTCTGCTACAGATGTTA	856
Db	806	TTCTGCGAGTTATCTGTCATTTATTGATTTCTTAGTGGAAAAAATACGAAACAGCAACGA	865
Qy	857	ACCAATATCTTTAGCATATACCAATGATTACTTTTGTCTGGCAGTCTGCTGCAAAAGTA	916
Db	866	ATGGATTTAATTTCAATTACACGAATATATACAAATTGATGGATGGCCAAAGGGAATG	925
Qy	917	AACCTTTTACTTTTAAGATGGACTGGATACAGAATAGTATGCCGGATCTTAACGGTATTG	976
Db	926	ATCCTCTTATATACTTTTGGACATCATACAAAATAGTGTATGATGATGATGCAATGGAGCTG	985
Qy	977	TCATTGTTGCTACAACATAGNACAGTTTACAGACAGTACACACTGCTGTCTACTTCTTACCAT	1036
Db	986	CCGTAGTGTGTTTCTACGAGAACAGTCACTTGATTTCTTCAACAGCAATTAACACATTTCCGT	1045
Qy	1037	TCAATCCAAGTGTGTGATAAAACCAAAACAAATTCGAATTTTTCGAACCTATTTCCAACTACTA	1096
Db	1046	TTGATCCACAGTTGATTAACCAAAACCACTTGAGTAAATAGAACCCATCCCTACTACCA	1105
Qy	1097	CCATCACAACTTCATATGTTGGTGGACTACTTCTCTATCTGACTAAGACTGCACCAATG	1156
Db	1106	CTATTACTACTTCATATGTTGGATTTCTTCTTCTACTTTCTACGAAGACTGCAACTATTG	1165

RESULT 12	
AF201684	
LOCUS	6897 bp DNA linear PLN 09-AUG-2000
DEFINITION	Candida albicans agglutinin-like protein Als7p (ALS7) gene, complete cds.
ACCESSION	AF201684
VERSION	AF201684.1 GI:9754770
KEYWORDS	
SOURCE	Candida albicans.
ORGANISM	Candida albicans
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE	1 (bases 1 to 6897)
AUTHORS	Hoyer, L.L. and Hecht, J.E.
TITLE	The ALS6 and ALS7 genes of Candida albicans


```
Qy 964 TCTAAGCGTATTGTCATTTGGTCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTC 1023
Db 970 GCTATTGGTCGAGACCTAGTTTCGTCACCATCCACAGTATAGGTTCCATTACCAAGAACT 1029
Qy 1024 ACTACTTTTACCATTCAAATCCAAAGTGTGATAAAACCAAAACAAATCGAATTTTGCACCT 1083
Db 1030 ACCACATTACCCCTTCATTTCCCGACTCCAGAAAACCAAAACAAATCTAGTCTTTAGAGCCC 1089
Qy 1084 ATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCTATCTGACTAAG 1143
Db 1090 ATACCCACCACTACCGTAAACAACITTCACACCATGCTTGTATATACTTGTATATACTAAG 1149
Qy 1144 ACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGGCCATATCATATACATACACA 1203
Db 1150 AAAGCCACCAATTTGGTGACACAGCTACTGTTTTCATTTGATGTTCCACACATACAGCTACT 1209
Qy 1204 ACTGTTACCAAGTGTGACAGGAGCAATCACTACCAACCAACTCGTACCAATCCCACT 1263
Db 1210 ACTTTGACCAATATTTGGCAAGAATCAAGTACAGGACAACCACTTACTTCGATGACATA 1269
Qy 1264 GATTCAATTGACACAGTGGTGTACAACTTCCACTGCCAAATCCAACTGTTTAGTACTACT 1323
Db 1270 GACTTGGTCGATCTGTCATTTGGAATTTCCATATCCCAATCCGACTATTATTAACAACA 1329
Qy 1324 GAATATGGTCTCAGTCTCTTGGCTACACCACTACAGTTACTGCTCTCCAGGTGGTACC 1383
Db 1330 CAATTTTGGTCAGGTAATATTTAACTACTGAGACACACAAGAACCACTCTCGGTACT 1389
Qy 1384 GATAGTGTATTATCAGAGGCCAACCAACCATACTGTCATCTACTACTGATATATTGGTCA 1443
Db 1390 GATAGTGTATCAGGAACCAACCAACCACTACTGTGACAAGCAGGAGTTTGGTCA 1449
Qy 1444 CAATCTTGTGCTACTACTACTGTTTACTGCTCTCCTCCAGGTGGTACTGACTCAGTAAT 1503
Db 1450 GAATCATTTGCCACTACTGAGACCATCAACCAACTATCCAGAAGGCAGTACTGACTCAGTCAT 1509
Qy 1504 ATCAGAGAACCAACCAATCCAACTGTCATACACCGAGTATTGGTCTCAATCTTGGCT 1563
Db 1510 GTTAGGGAACCAACCAACCACTGTCGACAAGCAGGAGTTTGGTCTGAGAATCATTTGCC 1569
Qy 1564 ACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCT 1623
Db 1570 ACTACTGAGACCATCAACCAACCGGTCAGAGGCAGTACTGACTCAGTCATTTAGGGAACCA 1629
Qy 1624 CCAACCAACTGTCAACCACTGATATTTGGTCCCAATTTAGGCAACCACTACT 1683
Db 1630 CATATCAACTGTGACAACCAACCGAGTTTGGTCTGAGAATCATTTGCCACTACTGAGACC 1689
Qy 1684 GTGACTGCTCTCCAGGAGGCAGTACTGACTCAGTAATTTATCAGAGAACCAACCACT 1743
Db 1690 ATCAACCAAGGTCAGAGAGGCAGTACTGACTCAGTCATTTAGGGAACCAACCAATCCAACT 1749
Qy 1744 GTCACCTACTGTAATCTGTCACAAATCATATGCAACCACTACCACTGTAACCTGCACCA 1803
Db 1750 GTGACAACCAAGGAGTTTGGTCAGATCATTTGCCACTACTGAGACCATCAACCAAGGT 1809
Qy 1804 CCAAGTGTGATGACACTGTTATCATTTAGAGAGCCACCAACCACTGTCACCTACTACT 1863
Db 1810 CCAGAAGCACTGACTCAGTCATTTAGGGAACCAACCAATCAACCTGTCACAACCACT 1869
Qy 1864 GAGTATTGGTCTCAATCTGTTGCTACTACCACTGTAACCTGGTCCCAAGTGGCACT 1923
Db 1870 GAGTTTGGTCAGATCATTTTGGCCTACTGAGACCATCAACCAAGGTCAGAGAGGCACT 1929
Qy 1924 GATAGTGTATCATTTAGGGAACCAACCAACCACTGTCACCACTACTGTAATCTGGTCT 1983
Db 1930 GATAGTGTATCATCAAGGAACCAACCAACCACTGTCGACTACTACCAAGTTTGGTCA 1989
Qy 1984 CAATCATATGCAACCACTACTACCAATTAACCGCTCCACTGTTGGAACCTGATACCGTTCT 2043
Db 1990 GAATCATTTGCCACTACTGAGACCAATCAACCACTATCCAGAAGGCAGTACTGACTCAGTCAT 2049
Qy 2044 ATCAGAGAGCCCAACCAACCACTACTGCTACTACTGTAATCTGCTCAATCATATGCT 2103
```

```
Db 2050 GTTAGGGAACCACTAATCCAACTGTGACAACACCGAGTTTGGTCAGAATCATTTGCC 2109
Qy 2104 ACAACCACTGTTTACTGCAACCACTGGTCAACCGATACCGTCTCTTATCAGAGACCA 2163
Db 2110 ACTACTGAGACCATCAACCAAGGTCGAGAGGCACTGACTCAGTCATTTGTTAGGAACCA 2169
Qy 2164 CCAACCACTGCTACTACTACTGTAATCTGCTCAATCATATGCTTACCAACCACTACT 2223
Db 2170 CACAATCAACTGTGACAACCAACCGAGTTTGGTCTGAGAATCATTTGCTACTGAGACC 2229
Qy 2224 GTTACTGCAACCAAGGTCGATACCACTGTTTATCATTTAGAGAGCCCAACCACTACT 2283
Db 2230 ATCAACCACTATCCAGAGGCACTGACTCAGTCATTTGTTAGAGAACCAACCACTACT 2289
Qy 2284 GTTACTACTGTAATTTGCTCAACATTTGCTCAACCACTGCTCAACCACTGCTCTCT 2343
Db 2290 GTACAACCAAGGAGTTTGGTCTGAGTCATTTGCTACTGAGACAGTCACCACTAT 2349
Qy 2344 CCAGGTGCTACTGACTGCTGATTTATCTATGAA 2376
Db 2350 CCAGAAGGCACTGACTCAGTCATTTGTTAGGAA 2382

RESULT 13
CANALS4S1
LOCUS
DEFINITION
Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
cds.
ACCESSION
AF024584
VERSION
AF024584.1 GI:3598677
KEYWORDS
1 of 2
SEGMENT
1
SOURCE
Candida albicans.
ORGANISM
Candida albicans.
Eukaryote: Fungi; Ascomycota; Saccharomycotina: Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS
Hoyer,L.L., Payne,T.L. and Hecht,J.E.
TITLE
Identification of Candida albicans ALS2 and ALS4 and localization
of als proteins to the fungal cell surface
J. Bacteriol. 180 (20), 5334-5343 (1998)
JOURNAL
98440424
MEDLINE
9765564
PUBMED
2 (bases 1 to 1407)
AUTHORS
Hoyer,L.L.
TITLE
Direct Submission
JOURNAL
Submitted (11-sep-1997) Veterinary Pathobiology, University of
Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
location/Qualifiers
source
1..1407
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="6"
/map="6C"
1..>1407
/gene="ALS4"
/note="agglutinin-like sequence"
/allele="ALS4-1"
1..>1407
/gene="ALS4"
/transl_start=1
/product="agglutinin-like protein"
/protein_id="AAC64239.1"
/db_xref="GI:3598680"
/translation="MLLQFLLSLCVSVATAKVTGVFNSFNLSLTWANAASYPYRGA
PFTWTAVIGNSLDGATASAGDTFTLDMPCVFKFTDQTSIDLVDAGRTYATCNLSAE
EPTTFSSVSCVTVTMTTADTKAIGTVPFPFSVSGSGSDVDLANSQCFAGINTVF
NDGTSISTTVDPEKSTVASSDRLLSRILPSLQAVNLFLPQCANGYTGTMGFT
AGTCATIDCSTVHVGIENGLDNWNPISSEFSYTKTCTSTSVLTVQNPVGYRPFV
DAYISATRVSSYTMQYTNIVACGAASVDDSFTHWRGYSNSQAGSNGITIVTRTV
```



```
CDS
1. >1407
/gene="ALS4"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC64241.1"
/db_xref="GI:3598685"
/translation="MLQLFLLSLCVSVATKAVITGVFNSFNSLTWANAASYPRGPA
EPTWATVGSGLDGTASAGDTFTLDMPCVFKETIDOTSIDLVDAGRTVATCNLSAE
EFTFSVSCTVTTTADTKAMGTLPFSVSGGSDVDLANSOCFTAGINTVTF
NDGDTISITVDFEKSTVASSDRILLRSLPSLSQAVSLFLPOECANGYSGTMGSEF
AGTGATDCSTVHVGSLNDWNYPISSEFSYTKTCTSYLVLYQNPVPGYRPFV
DAYVSATRVSSYTMQVNIYACVGAASVDDSEFTHWGYSNSQAGSNGITVIVTTRV
TDSSTAVTLFNSDDTKTIEIQLPIPTTTTTSYGVNTTSYKTAIPIGTATVI
VDYVHTTTLTSEMGTITTTTTRINPTDSIDTVVQVPSNPPTVTITTEYNSQSYAT
TTVTAPPGGTGSVIREP"
1300..1407
repeat_region
/note="first of multiple copies of 108bp imperfect repeat"
/rpt_type=tandem
BASE COUNT 389 a 297 c 257 g 464 t
ORIGIN
Query Match 22.1%; Score 836.4; DB 8; Length 1407;
Best Local Similarity 74.7%; Pred. No. 4e-141;
Matches 1050; Conservative 0; Mismatches 356; Indels 0; Gaps 0;
QY 1 ATGCTTCAACAATTTACATTTGTTATTCCTATATTGTCATTTGCAATGCGAAGTGCAAGACAAATC 60
DB 1 ATGCTTTTACAATTTTGTGCTAAGCCTCTGTGTATCAGTTGCTACGGCAAAAGTTATT 60
QY 61 ACTGGTGTTTTGATAGTTTAAATCAATTAACCTGGTCCCAATGCTGCTAATTAATGCTTTC 120
DB 61 ACAGGTGTTTTCATAGTTTAAATTCGTTAACTTGGGCCAATGCTGCTTATTCACATAT 120
QY 121 AAAGGCCAGGATACCAACTTGAATGCTGTTTGGTGGTCTTAGATGGTACAGT 180
DB 121 AGAGTCCAGTACTCTCTTGGACGCGTGAATAGATGGTCTTTAGATGGAGCTACT 180
QY 181 GCCAATCAGGGGATACATTACATATGAATATGCATGTGTGTTTAAATATACACTTCA 240
DB 181 GCTAGTCTGGTGACATTTACCGTTAGACATGCTTGTGTTTCAAAATTTATTAAGTAT 240
QY 241 CAACATCTGTTGATTTAACTGCGGATGGTGTAAATATGCTACTGTCATTTTATTTCT 300
DB 241 CAACAGTCAATTTAGTTAGTTGCTGATGGTGGTACTTATGCTACTTGTAAATTTGAAC 300
QY 301 GGTGAAGAATTCACAACTTTTCTACATTAACATGTACTGTGAACGACGCTTTCAAAATCA 360
DB 301 GCCGAGAGTTTACTACTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACTGCT 360
QY 361 TCCATTAAGGCATTTGGTACAGTTTACTTTACCAATTCATTCATTAATGTTGGTGAACAGGT 420
DB 361 GACACCAAGGCATGGGAACGTGAACATACCTTTCTCATTCACTGTTGGGGGATCAGGT 420
QY 421 TCATCACTGATTTGGAGAGTCTTAATGTTTACTGCTGCTACCAATACAGTCACATTT 480
DB 421 TCAGATGTTGATTTGGCAAAATTCCTCAATGTTTACTGCAGGAATCAATACAGTTTACTTT 480
QY 481 AATGATGCTAATAAGATATCTCAATTTGATGTTGAGTTTGAAGAATCAACCGTTGATCCA 540
DB 481 AATGATGGTACACTAGCATTTCCACACAGTTGATTTTGAANAATCAACCGTGGCCCTCC 540
QY 541 AGTGCATTTTGTATGCTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACCTCTTTTT 600
DB 541 AGCGATGCTATCTTGTGTCAAGAATTTTACCAGCTTTTCAACAGCAGTAAGTCTTTTT 600
QY 601 GTGGCACCACCAATGTGAATGGTTTACACATCTGGTACATGGGTTCTCCAGTAGTAAC 660
DB 601 CTTCCCCAAGAATGTGCAATGGTTATCTCTGGTCAATGGGATTTTCAACTGCTGGT 660
QY 661 GGTGACGTTGCTATTGCTGCTCAAAATTCATATTTGGTATCACAAAAGGATTAATGAT 720
DB 661 ACTGGTCTACTAGATTGTTCCACAGTTCATGCTCGGGATATCAATGCGTTGAATGAT 720
```

```
QY 721 TGAATTATCGGTTTCATCTGAATCATTTAGTTACACTAAACTTGTACATCTAATGGA 780
DB 721 TGAATTATCCAATTTTCCATGGAATCTTTTCTTACACAAGACCTGTACATCAACAGT 780
QY 781 ATTGAGATTAAATATCAAAATGATGCTGCTGTTATCGTCCATTTATTTATGCTTATTT 840
DB 781 GTTTAGTAACTTATCAAGAATGTTTCCCTGCGGGATATCGTCCATTTGTCGATTTATGTT 840
QY 841 TCTGCTACAGATGTTACCAATATATCTTTAGCATATATACCAATATACCAATGATTTGCTGGC 900
DB 841 TCTGCAACACAGAGTCAAGCTCATATACCATGCAATACACTAATATATATGTTGTTGGC 900
QY 901 AGTCTCTGCAAAATCAACCTTTTCCATTTTAAAGATGGAGTGGATACAAAGATAGTGAATGCC 960
DB 901 CGGCTTCTGTGTGATGACTCATTTACTTACTTACTTGGCGGGATATAGTAATAGTCAAGCT 960
QY 961 GGATCTAACGGTATTGTCATTTGTTGCTTACAACTAGACAGTGTACAGACAGTACCACTGCT 1020
DB 961 GGTCTTAATGGTATTACCAATTTGTTGTAACAACTAGACAGTGTACAGACAGTACCACTGCT 1020
QY 1021 GTCACTACTTTTACCATTCAATCCAAAGTGTGATAAAACCAAAACAAATCGAAATTTTCAA 1080
DB 1021 GTGACTACTTTTACCATTCAATTCGATACACTGACAAAACCAAAACAAATCGAAATTTTCAA 1080
QY 1081 CCTATTTCCAACTTACCATTACCACTTCAATTTATGTTGGTGTGACTACTTCTTATCTGACT 1140
DB 1081 CCTATTTCCAACTTACCATTACCACTTCAATTTATGTTGGTGTGACTACTTCTTATCTGACT 1140
QY 1141 AAGACTGACCAATTTGTTGTAACAGTACTTGTATTTGTTGATGTGCTATATCATATACCTACC 1200
DB 1141 AAGACTGACCAATTTGTTGTAACAGTACTTGTATTTGTTGATGTGCTATATCATATACCTACC 1200
QY 1201 ACAACTGTTTACCAGTGAATGGAGCAAGCAATCAATCACTACCACCAACTCGTACCAATCCA 1260
DB 1201 ACAACTGTTTACCAGTGAATGGAGCAAGCAATCAATCACTACCACCAACTCGTACCAATCCA 1260
QY 1261 ACTGATTTCAATTTGACAGTGGTGGTAACTTCCAACTTCCAACTGCAAACTGTTAGTACT 1320
DB 1261 ACTGATTTCAATTTGACAGTGGTGGTAACTTCCAACTTCCAACTGCAAACTGTTAGTACT 1320
QY 1321 ACTGAATTTGGTCTCAGTCTTGGTACAACTACAGTGTGCTTCTCTCCAGGTTGGT 1380
DB 1321 ACCGAGTATTTGGTCTCAGTCTTGGTACAACTACAGTGTGCTTCTCTCCAGGTTGGT 1380
QY 1381 ACCGATACTGTGATTTATCAGAGAGCC 1406
DB 1381 ACTGACTCAGTGTATTCAGAGACC 1406
```

RESULT 15
AF229989S1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1404 bp DNA linear PLN 02-FEB-2001
Candida albicans agglutinin-like protein (ALS9) gene, partial cds.

AF229989.1 GI:12656144

1 of 2
Candida albicans.

Candida albicans.

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

Hoyer, L.L., Hecht, J.E. and Mirus, K.A.

The ALS9 gene of Candida albicans

Unpublished

2 (bases 1 to 1404)

Hoyer, L.L.

Direct Submission

Submitted (31-JAN-2000) Veterinary Pathobiology, University of

Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

Location/Qualifiers

1..1404


```
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="6"
/map="60"
1. >1404
/genes="ALS9"
/codon_start=1
/product="agglutinin-like protein"
/protein_id="AAK00764.1"
/db_xref="GI:12656146"
/translation="MLPQIFLLISLTSTAKTIRGVNFSDFSLTWRSVEYAKGPE
TPTWAVLWLSNSTADPGDFTLILPCVFKFITQTSVDLTADGVSYATCDFNAGE
EFTTSSLSCTVNSVSARVGVKLPITENVGKTSSVLDLADSKCFTAGKNVTF
MDGNKISTVDVDFDASVPSGYITSSRIPLSLKSLFVVPQCEGYTSGINGVFA
SNGAIDSGNVLIGISKGLNDWNPVSSSESYTKTSTISITVFQNPVAGRFVD
ATISENDKYLITANEYTCENGVDPFLLTWGYNKSEADSDGVIVVVITRTV
DSTAVTLTPNPSPVDKTEIILQPIPTTTTTSYIGISTSYETLTIGGTATVIV
DTPYHITATVNFMTGSIITTTTNTPTGSDIVIVQIPSPDPTTTITEFWSEFAST
TTITNPDPGTSNVIKEP"
-repeat_region
1297..>1404
/notes="first of a series of 108 bp repeat units"
/rpt_type=tandem
BASE COUNT 413 a 286 c 253 g 452 t
ORIGIN
Query Match 21.3%; Score 805.4; DB 8; Length 1404;
Best Local Similarity 73.8%; Pred. No. 1.6e-135;
Matches 1038; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 1 ATGCTTCAACAATTACATTTGTTATTCCTATATTTGTCATTTGCAATTCGAAGTGCAGAACACAATC 60
Db 1 ATGCTTCCACAATTCATATTTGTTATTCATATCTTCACAGTGCAGTGCAGTGCAGAACTATTT 60
QY 61 ACTGGTGTGTTTGAATAGTTTAAATTCATTAACCTTGGTCCAAAGTCTGCTCAATTAATGCTTTC 120
Db 61 ACTGGTGTGTTTCAATAGTTTGAATTCATTCACATGGACTAGATCCGCTTGAATATGCTTAC 120
QY 121 AAAGGCCAGGATACCAACTTGGATGCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180
Db 121 AAAGGCCAGAGACTCCAACCTTGGATGCAATTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 181 GCCATCCAGGGGATACATTCACATTCGAATATGCCATGCTGTGTTTAAATATACACTTCA 240
Db 181 GCTGACCCAGGACACATTCACCTTGATTTGCTTGTGCTATTTAAATTTTAACTACC 240
QY 241 CAAACATCTGTTGATTTAAGTCCGATGGTGGTAAATATGCTACTTGTCAATTTATCT 300
Db 241 CAAACATCTGTTGATTTGATGCTGATGGTGGTAAATATGCTACTTGTGACTTTAATGCT 300
QY 301 GGTGAAGAATTCACAACCTTTTCTACATTAACATGACTCTGAACGACGCTTTTGAATCA 360
Db 301 GGTGAAGAATTCAGACATTTTCTTCCATATGACTGTGAACAGTGTTCGGTATCA 360
QY 361 TCATTAAGGCATTTGGTACAGTTACTTTACCAATTCATTCATTAATGTTGGTGGAAACAGGT 420
Db 361 TATGCTAGGTTTCTGTAAGGTCAAATTCGCCATTCATTCATTAATGTTAGTGGAAACAGGT 420
QY 421 TCATCAACTGATTTGGAAGATCTTAATGTTTACTGCTGGTACCAATACAGTCACATTT 480
Db 421 TCTTCAGTTGATTTGGCAGATTCCAATGTTTACTGCCGGAACAAACACTGTGACTTTC 480
QY 481 AATGATGGTGATTAAGATATCTCAATTTGATGTTGAGTTTCAAAAGTCAACCGTTGATCCA 540
Db 481 ATGGATGGCGGATCAAAAGATTTTACCACTGTGTTGATTTGACGCGTCTCCAGTATCACCC 540
QY 541 AGTGCATATTTGTTATGCTTCCAGATTTATGCCAAGTCTCAATTAAGGTCAACACTCTTTT 600
Db 541 AGTGGTTATATACAAGCTCACCAATATTTCCTAGTCTCAATTAACCTTCAAGTCTTTTC 600
QY 601 GTGGCACCAATGTGAAATGGTTACACATCTGGTACAAATGGGGTTCTCCAGTAGTAAC 660
Db 601 GTGGTCCACAAATGTGAGAATGGTTACACATCTGGTATAATGGGATTTGTAGCTAGTAAC 660
QY 661 GGTGACGTTGCTATTCCTCAAAATATTCATATTTGGTATACAAAGAGATTAAATGAT 720
Db 661 GGT---GCTACTATTTGATTCCTCAAAATGCAATATAGGAATATCAAAAGGTTTAAATGAT 717
QY 721 TGAATATATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
Db 721 TGAATATATCCAGTAAAGTTTCAGAAATCATTTCTTACACGAAACTTGTAGTCAACAGT 777
QY 781 ATTCAGATTAATATCAAAATGTACCTGCTGGTTATCTGCTCCATTTATGATGCTTATTT 840
Db 778 ATTACAGTTGAATTTCAAAATGTCTGCTGGTATCGGCCCTTTTGTGATGATATATTT 837
QY 841 TCTGCTACAGATGTTAAACCAATATCTTTAGCATATACCAATCATATATCTTCTGCTGCG 900
Db 838 TCTGCAGAAATATTCGTAATATACCTTTCAGCTAGCAANTAGTATATCTTCTGTAAT 897
QY 901 AGTCTGCTGCAAAAGTAAACCTTTTCACTTTAAGATGACCTGGATACAGAAATAGTATGCC 960
Db 898 GGCAATACTGCTGGTTGATCCATTTACTTTAAACATGGTGGGGGTATAAAAACTCTGAAGCA 957
QY 961 GGATCTAAACGGTATTTGCTATTTGCTACAACTAGAACAGTTACACAGACTACACTGCT 1020
Db 958 GACTCTGACGGGATGCTGATCTGCTAGTTACAAACCAAGTGTACACAGACTACACAGCT 1017
QY 1021 GTCACACTTTTACCATTCATTCACAACTTCATATGTTGGTGTGACTACTTCTCTACTGACT 1140
Db 1018 GTGACTACTTTTACCTTTCAATCCAGTGTGATTAACCACTTCATATATTTGCTATTTCC 1137
QY 1141 AGACTGACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTGCTATCATATCTACTACC 1200
Db 1138 TTAACCGGAACAATTTGGTGTGCTGACAGTCAATTTGCTGATACACCTTATCATATCACT 1197
QY 1201 ACAACTGTTTACCAAGTGAATGGACAGGAACAATCACTACCAACCAACTCTGACCAATCCA 1260
Db 1198 GCCACTGTTTACAATTTCTGGACTGGTCAATTTACAACCTACCCTACTTACTATATCCC 1257
QY 1261 ACTGATTTCAATTCACAGTGGTGGTACAAAGTTCCACTGCCAAATCCAATGTTAGTACT 1320
Db 1258 ACTGGTTCCATACACACTGTTATTTGCAAAATTTCCACTGCTGCTGATCCAACCTACAACTATA 1317
QY 1321 ACTGAATATTTGGTCTCAGTCCCTTTGCTACAAACCACTACAGTTACTGCTCTCCAGGTGGT 1380
Db 1318 ACTGAATTTTGGTCTGAATCATTTGCTAGTACTACCACTACCACTACCACTGAGGTT 1377
QY 1381 ACCGATACTGTGATTTATCAGAGAGCCA 1407
Db 1378 ACTAATAGTGTGATCATCAAGAAGCA 1404
```

Search completed: April 29, 2003, 19:18:22
Job time : 6516 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search; using sw model

Run on: April 29, 2003, 15:36:18 ; Search time 516 seconds
(without alignments)
16523.383 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgcttcacaaattacatt.....tgtcttgcatttagtga 3786

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.2	9.5	424	17 AAT29067	Probe for Candida
2	108	2.9	108	17 AAT32337	Probe for Candida
3	108	2.9	108	17 AAT32330	Probe for Candida
4	108	2.9	108	17 AAT32331	Probe for Candida
5	108	2.9	108	17 AAT32332	Probe for Candida
6	108	2.9	108	17 AAT32334	Probe for Candida
7	108	2.9	108	17 AAT32336	Probe for Candida
8	108	2.9	108	17 AAT29063	Probe for Candida
9	106.4	2.8	108	17 AAT32333	Probe for Candida

10	106.4	2.8	108	17 AAT32335	Probe for Candida
11	99	2.6	99	17 AAT32338	Probe for Candida
12	94.2	2.5	8201	21 AAA88664	Human dentin sialo
13	94.2	2.5	8201	24 ABQ73537	Human dentin sialo
14	87	2.3	5511	21 AAA61847	Cryptosporidium pa
15	87	2.3	7334	21 AAT04776	C parvum Gp900 gen
16	87	2.3	7334	21 AAA61846	Cryptosporidium pa
17	87	2.3	7334	24 AAT04775	C parvum Gp900 gen
18	86.2	2.3	5163	19 AAV20700	Cryptosporidium pa
19	86.2	2.3	5163	21 AAA61849	ORF encoding a por
20	86.2	2.3	5163	24 AAT04778	C parvum Gp900 gen
21	86.2	2.3	5318	19 AAV20701	Cryptosporidium pa
22	86.2	2.3	5318	21 AAA61848	DNA encoding a por
23	86.2	2.3	5318	24 AAT04777	C parvum Gp900 gen
24	85.2	2.3	108	17 AAT29064	Probe for Candida
25	82.4	2.2	100	15 AAQ62588	Candida albicans-s
26	82.4	2.2	100	15 AAQ62589	Candida albicans-s
27	81.4	2.2	7758	24 ABL33102	Human immune syste
28	76	2.0	100	15 AAQ62589	Candida albicans-s
29	74.4	2.0	267156	24 ABL68560	Kidney cancer rela
30	68.4	1.8	40875	18 AAT80043	Insert from cosmid
31	68	1.8	100	15 AAQ62592	Candida albicans-s
32	68	1.8	7758	24 ABL33103	Human immune syste
33	67.6	1.8	2336	23 ABL25662	Drosophila melanog
34	64.8	1.7	2790	7 AAN60473	Sequence encoding
35	62.6	1.7	1368	10 AAN91235	DNA sequence of pr
36	62.2	1.6	2014	24 ABA90791	Bacillus anthracis
37	58	1.5	3168	20 AAX36552	C. albicans Rbt1 c
38	58	1.5	3983	23 ABL15838	Drosophila melanog
39	57.4	1.5	2335	23 ABL25476	Drosophila melanog
40	57.2	1.5	100	15 AAQ62593	Candida albicans-s
41	57.2	1.5	3399	17 AAT05868	Chicken leucocytoz
42	57.2	1.5	32392	24 ABL56203	AmEPV genome fragm
43	57.2	1.5	50000	24 ABL5643	AmEPV genome fragm
44	57	1.5	100	15 AAQ62591	Candida albicans-s
45	56.6	1.5	14987	24 ABL32630	Human immune syste

ALIGNMENTS

RESULT 1

AAT29067
ID AAT29067 standard; DNA; 424 BP.

XX AAT29067;

XX DT 28-NOV-1996 (first entry)

XX DE Probe for Candida albicans and Candida stellatoideis.

XX KW Probe: primer; detection; identification; Candida albicans;

XX KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

XX KW lymph fluid; skin; soft tissue; ss.

XX OS Synthetic.

XX PN WO9618745-A1.

XX PD 20-JUN-1996.

XX PF 08-DEC-1995; 95WO-US16153.

XX PR 16-DEC-1994; 94US-0357962.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Hoyer LL, Livi GP, Shatzman A;

XX DR WPI; 1996-300661/30.

XX PT C. albicans and C. stellatoideis specific probes and primers - for

PT specific detection of Candida infection

DE PRO

CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.
 XX
 SQ Sequence 108 BP; 29 A; 29 C; 19 G; 31 T; 0 other;

Query Match 2.9%; Score 108; DB 17; Length 108;
 Best Local Similarity 100.0%; Pred. No. 2.4e-15;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTGTCTACAACCACTACA 1359

DB 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTGTCTACAACCACTACA 60

QY 1360 GTTACTGCTCCAGGTGGTACCGATAGTACTGATTATCAGAGAGCCA 1407

DB 61 GTTACTGCTCCAGGTGGTACCGATAGTACTGATTATCAGAGAGCCA 108

RESULT 9

AAT32333
 ID AAT32333 standard; DNA; 108 BP.

XX - AAT32333;

DT 29-NOV-1996 (first entry)

DE Probe for Candida albicans and Candida stellatoideis.

KW Probe; primer; detection; Identification; Candida albicans;
 KW Candida stellatoideis; sputum; bronchial washings; blood; milk;
 KW lymph fluid; skin; soft tissue; ss.

OS Synthetic.

PN WO9618745-A1.

PD 20-JUN-1996.

PF 08-DEC-1995; 95WO-US16153.

PR 16-DEC-1994; 94US-0357962.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX C. albicans and C. stellatoideis specific probes and primers - for
 PT specific detection of Candida infection

PS Claim 1; Figure 5A; 33pp; English.

XX Five synthetic sequences (AAT29063-67) are used as probes to detect
 CC the presence of Candida albicans and Candida stellatoideis in
 CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.

XX Sequence 108 BP; 34 A; 33 C; 16 G; 25 T; 0 other;

Query Match 2.8%; Score 106.4; DB 17; Length 108;
 Best Local Similarity 99.1%; Pred. No. 5.5e-15;
 Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1732 CCAACACACTGTACTACTGAATCTGCTCAATCATATGCGCCACTACTACC 1791

DB 1 CCAACACACTGTACTACTGAATCTGCTCAATCATATGCGCCACTACTACC 60

RESULT 10

AAT32335
 ID AAT32335 standard; DNA; 108 BP.

XX AAT32335;

DT 29-NOV-1996 (first entry)

DE Probe for Candida albicans and Candida stellatoideis.

KW Probe; primer; detection; Identification; Candida albicans;
 KW Candida stellatoideis; sputum; bronchial washings; blood; milk;
 KW lymph fluid; skin; soft tissue; ss.

OS Synthetic.

PN WO9618745-A1.

PD 20-JUN-1996.

PF 08-DEC-1995; 95WO-US16153.

PR 16-DEC-1994; 94US-0357962.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX C. albicans and C. stellatoideis specific probes and primers - for
 PT specific detection of Candida infection

PS Claim 1; Figure 5A; 33pp; English.

XX Five synthetic sequences (AAT29063-67) are used as probes to detect
 CC the presence of Candida albicans and Candida stellatoideis in
 CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.

XX Sequence 108 BP; 33 A; 37 C; 14 G; 24 T; 0 other;

Query Match 2.8%; Score 106.4; DB 17; Length 108;
 Best Local Similarity 99.1%; Pred. No. 5.5e-15;
 Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1948 CCAACCACTGTCACTACTGCTCAATCATATGCGCCACTACTACC 2007

DB 1 CCAACCACTGTCACTACTGCTCAATCATATGCGCCACTACTACC 60

QY 2008 ATTACCGCTCCAGTGGTGAACCTGATACCGTCTTATCAGAGAGCCA 2055

DB 61 ATTACCGCTCCAGTGGTGAACCTGATACCGTCTTATCAGAGAGCCA 108

RESULT 11

AAT32338
 ID AAT32338 standard; DNA; 99 BP.

XX AAT32338;

XX

```

DT 29-NOV-1996 (first entry)
XX Probe for Candida albicans and Candida stellatoideis.
DE
XX
XX Probe: primer: detection; identification; Candida albicans;
KW Candida stellatoideis; sputum; bronchial washings; blood; milk;
KW lymph fluid; skin; soft tissue; ss.
XX Synthetic.
OS
XX WO9618745-A1.
PN
XX 20-JUN-1996.
PD
XX 08-DEC-1995; 95WO-US16153.
XX
XX 16-DEC-1994; 94US-0357962.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
PA
XX Hoyer LL, Livi GP, Shatzman A;
PI
XX WPI; 1996-300661/30.
DR
XX
XX C. albicans and C. stellatoideis specific probes and primers - for
PT specific detection of Candida infection
PT
XX Claim 1; Figure 5A; 33pp; English.
XX
XX Five synthetic sequences (AAT29063-67) are used as probes to detect
CC the presence of Candida albicans and Candida stellatoideis in
CC biological fluids e.g. sputum, bronchial washings, blood, milk and
CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
CC method of detection may also comprise using two of the probe
CC sequences (AAT29065, AAT29066) as amplification primers and then
CC contacting one of the probe sequences with the amplified product and
CC detecting hybridisation.
XX
XX Sequence 99 BP; 29 A; 29 C; 14 G; 27 T; 0 other;
SQ
Query Match 2.6%; Score 99; DB 17; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.7e-13;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2272 CCAATCCACAGTTACTACTGTAATATTTGGTCACAATATTGGCCACACACCACA 2331
Db 1 CCAATCCACAGTTACTACTGTAATATTTGGTCACAATATTGGCCACACACCACA 60

Qy 2332 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
Db 61 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 99

RESULT 12
AA88864/C
ID AA88864 standard; DNA: 8201 BP.
XX
XX AA88864;
XX
XX 19-FEB-2001 (first entry)
DT
XX Human dentin sialophosphoprotein gene.
DE
XX Dentin sialophosphoprotein; DSPP; human; SIBLINGS;
KW integrin-binding ligand; diagnosis; antiinflammatory; therapy;
KW osteoporosis; chromosome 4; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 2387..7896
FT CDS /*tag= a
FT /*note= "contains exons"
FT

```

```

FT exon 2387..2437
FT /*tag= b
FT intron 2438..3576
FT /*tag= c
FT exon 3577..3660
FT /*tag= d
FT intron 3661..3793
FT /*tag= e
FT exon 3794..4780
FT /*tag= f
FT intron 4781..5256
FT /*tag= g
FT exon 5257..7896
FT /*tag= h
XX
XX WO200062065-A1.
PN
XX 19-OCT-2000.
XX
XX 07-APR-2000; 2000WO-US09349.
XX
XX 09-APR-1999; 99US-0128468.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Fisher LW, Fedarko NS, Young MF;
PI
XX WPI; 2000-679515/66.
DR
XX P-PSDB; AAB19772.
XX
XX Detecting small integrin-binding ligand N-linked glycoproteins for
PT detection of a tumor or protection against a complement mediated immune
PT response, comprises detection where Factor H is not an inhibitor -
XX
XX Disclosure; Page 98-100; 110pp; English.
XX
XX The present sequence is that of DNA encoding human dentin
CC sialophosphoprotein (DSPP, see AAB19772), a member of the small
CC integrin binding ligand, N-linked glycoproteins (SIBLINGS) family.
CC The invention provides methods and compositions for exploiting the
CC discovery that members of the SIBLINGS family bind to complement
CC Factor H, conferring resistance to complement mediated lysis. A
CC claimed method of conferring protection against a complement
CC mediated immune response involves providing a reservoir or other
CC supply in the subject's body so that a SIBLINGS protein can be
CC dispersed to interfere with complement mediated lysis and
CC inflammation. This protects cells that are grafted onto foreign
CC tissue or bone marrow cells introduced into a foreign host. The
CC SIBLINGS protein can be BSP, OPN, DMP1 or DSPP. A method of
CC detecting a SIBLINGS protein in a sample from a subject suspected
CC of having abnormal bone turnover, especially osteoporosis, is also
CC claimed.
XX
XX Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
SQ
Query Match 2.5%; Score 94.2; DB 21; Length 8201;
Best Local Similarity 41.7%; Pred. No. 1.6e-11;
Matches 588; Conservative 0; Mismatches 823; Indels 0; Gaps 0;

Qy 952 AGTCATCGCGATCTAACGGTATTGTCATGCTCTACACTAGACAGTTACACAGT 1011
Db 7290 ACTGCTGTCTACTATTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7231
Qy 1012 ACCACTGCTGTCTACTACTTTTACCATTCAATCCAAGTGTGTGATAAACAACAAATCGAA 1071
Db 7230 ATCGCTGCTGCTACTGCTGTGTGCTGCTGCTACTACTATTACTGCTATCATCTGCTCACT 7171
Qy 1072 ATTTTGCAACCTATTTCACCAACACTACCACACTTCATATGTTGGTGACTACTTCC 1131
Db 7170 GCTGTCACTGCTATCACTGCTGTGTGCTGCTATCACTGCTGCTGCTGCTATC 7111
Qy 1132 TATCTGACTAAGACTGCGACCAATTTGGTGAACACCTACTGTTATTGTTGATGTCCTAT 1191

```

[illegible]

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 29, 2003, 16:36:22 : Search time 90 Seconds
(without alignments)
12900.869 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgtctcaacaattacatt.....tgctctgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCITUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1069.4	28.2	1071	1	US-08-357-962-1
2	1069.4	28.2	1071	1	US-08-878-106-1
3	348.2	9.2	425	1	US-08-357-962-5
4	348.2	9.2	425	1	US-08-878-106-5
5	87	2.3	5511	3	US-08-928-361B-2
6	86.2	2.3	7334	3	US-08-928-361B-1
7	86.2	2.3	5163	3	US-08-700-651-1
8	86.2	2.3	5163	3	US-08-928-361B-4
9	86.2	2.3	5318	3	US-08-700-651-2
10	86.2	2.3	5318	3	US-08-928-361B-3
11	85.2	2.3	108	1	US-08-357-962-2
12	85.2	2.3	108	1	US-08-878-106-2
13	82.4	2.2	100	1	US-08-145-705A-32
14	82.4	2.2	100	1	US-08-145-705A-34
15	76	2.0	100	1	US-08-145-705A-33
16	68	1.8	100	1	US-08-145-705A-36
17	58	1.5	3168	4	US-09-165-239A-3
18	57.2	1.5	100	1	US-08-145-705A-37
19	57	1.5	100	1	US-08-145-705A-35
20	54.2	1.4	3489	2	US-08-728-329A-1
21	54.2	1.4	3489	4	US-09-298-568-1
22	54.2	1.4	32207	2	US-08-770-379-20
23	54.2	1.4	32207	4	US-08-757-669A-20
24	54.2	1.4	32207	4	US-09-230-371A-20
25	50.2	1.3	4197	2	US-08-682-517-7
26	50.2	1.3	4197	2	US-08-682-517-8
27	49	1.3	3337	1	US-08-072-610-1

c	28	49	1.3	3337	2	US-08-719-822B-1	Sequence 1, Appl
c	29	49	1.3	3337	4	US-09-092-458-1	Sequence 1, Appl
c	30	48.8	1.3	3183	4	US-08-911-393-1	Sequence 1, Appl
	31	47.8	1.3	2824	4	US-07-757-022B-13	Sequence 13, Appl
	32	47.8	1.3	3066	4	US-07-757-022B-83	Sequence 83, Appl
	33	47.8	1.3	3117	4	US-07-757-022B-73	Sequence 73, Appl
	34	47.8	1.3	3148	4	US-07-757-022B-57	Sequence 57, Appl
	35	47.8	1.3	3420	4	US-07-757-022B-103	Sequence 103, Appl
	36	47.8	1.3	3813	4	US-07-757-022B-43	Sequence 43, Appl
	37	47.8	1.3	3936	4	US-07-757-022B-41	Sequence 41, Appl
	38	47.8	1.3	3942	4	US-07-757-022B-141	Sequence 141, Appl
	39	47.8	1.3	3945	4	US-07-757-022B-49	Sequence 49, Appl
	40	47.8	1.3	3963	4	US-07-757-022B-45	Sequence 45, Appl
	41	47.8	1.3	3963	4	US-07-757-022B-59	Sequence 59, Appl
	42	47.8	1.3	4065	4	US-07-757-022B-47	Sequence 47, Appl
	43	47.8	1.3	4086	4	US-07-757-022B-39	Sequence 39, Appl
	44	47.8	1.3	4092	4	US-07-757-022B-51	Sequence 51, Appl
	45	47.8	1.3	4215	4	US-07-757-022B-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-08-357-962-1
; Sequence 1, Application US/08357962
; Patent No. 5668263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/357,962
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1071 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-357-962-1

Query Match 28.2%; Score 1069.4; DB 1; Length 1071;
Best Local Similarity 99.9%; Pred. No. 5.8e-252;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCAACTGTAGTACTACTGAATATTGGTCTCAGTCCCTTTGTACAAACCACTACA 1359
DB 1 CCAATCAACTGTAGTACTACTGAATATTGGTCTCAGTCCCTTTGTACAAACCACTACA 60
QY 1360 GTTACTGCTCCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
DB 61 GTTACTGCTCCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
QY 1420 GTCACTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTGTACTGCTCCT 1479
DB 121 GTCACTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTGTACTGCTCCT 180
QY 1480 CCAGTGGTACTGACTCAGTAATTTATCAGAGAACCCCAATCACTGTCACCTACCAACC 1539
DB 181 CCAGTGGTACTGACTCAGTAATTTATCAGAGAACCCCAATCACTGTCACCTACCAACC 240
QY 1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCCTCCAGGTGGTACT 1599
DB 241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACTGCTCCTCCAGGTGGTACT 300
QY 1600 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 1659
DB 301 GACTCAGTAATTTATCAGAGAACCTCCAAACCACTGTCACCACTGTAATATTGGTCC 360
QY 1660 CAATCTTACGCAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 1719
DB 361 CAATCTTACGCAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 420
QY 1720 ATCAGAGAACCCAAACCACTGTCACCTACTACTGTAATTTGTCACCAATCATATGCC 1779
DB 421 ATCAGAGAACCCAAACCACTGTCACCTACTACTGTAATTTGTCACCAATCATATGCC 480
QY 1780 ACCACTACCACTGTAATTTGTCACCACTGTCACCTACTACTGTAATTTGTCACCACT 1839
DB 481 ACCACTACCACTGTAATTTGTCACCACTGTCACCTACTACTGTAATTTGTCACCACT 540
QY 1840 CCAACCACTGTCACCTACTACTGTAATTTGTCACCTACTACTGTAATTTGTCACCACT 1899
DB 541 CCAACCACTGTCACCTACTACTGTAATTTGTCACCTACTACTGTAATTTGTCACCACT 600
QY 1900 GTAAGTGGTCCAAAGTGGCACTGATGTTATCATTAGGGAACCAACCAACCACT 1959
DB 601 GTAAGTGGTCCAAAGTGGCACTGATGTTATCATTAGGGAACCAACCAACCACT 660
QY 1960 GTCACCACTACTGTAATTTGTCACCACTGTCACCTACTACTGTAATTTGTCACCACT 2019
DB 661 GTCACCACTACTGTAATTTGTCACCACTGTCACCTACTACTGTAATTTGTCACCACT 720
QY 2020 CCTGGTAACTGATACCGTCTTATCAGAGAGCCCAACCACTACTGTCACCTACT 2079
DB 721 CCTGGTAACTGATACCGTCTTATCAGAGAGCCCAACCACTACTGTCACCTACT 780
QY 2080 GAATPACTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCACCACTGTTGAAACC 2139
DB 781 GAATPACTGGTCTCAATCATATGCTTACCAACCACTGTTACTGCACCACTGTTGAAACC 840
QY 2140 GATACCGTCTTATCAGAGAGCCCAACCACTACTGTCACCTACTGTAATTTGTCACCT 2199
DB 841 GATACCGTCTTATCAGAGAGCCCAACCACTACTGTCACCTACTGTAATTTGTCACCT 900
QY 2200 CAATCATATGTCACCAACCACTGTTACTGCACCACTGTTGTCACCACTGTTGTAATC 2259
DB 901 CAATCATATGTCACCAACCACTGTTACTGCACCACTGTTGTCACCACTGTTGTAATC 960
QY 2260 ATTAGAGAGCCCAACCACTGTTACTGTAATTTGTCACCACTGTTGTAATTTGTCACCACT 2319
DB 961 ATTAGAGAGCCCAACCACTGTTACTGTAATTTGTCACCACTGTTGTAATTTGTCACCACT 1020
QY 2320 ACAACCACTGTAATTTGTCACCACTGTTGTCACCACTGTTGTAATTTGTCACCACT 2370

DB 1021 ACAACCACTGTAATTTGTCACCACTGTTGTCACCACTGTTGTAATTTGTCACCACT 1071

RESULT 2

US-08-878-106-1
; Sequence 1, Application US/08878106
; Patent No. 5817466
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,106
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/357,962
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1071 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-878-106-1

Query Match 28.2%; Score 1069.4; DB 1; Length 1071;
Best Local Similarity 99.9%; Pred. No. 5.8e-252;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCAACTGTAGTACTACTGAATATTGGTCTCAGTCCCTTTGTACAAACCACTACA 1359
DB 1 CCAATCAACTGTAGTACTACTGAATATTGGTCTCAGTCCCTTTGTACAAACCACTACA 60
QY 1360 GTTACTGCTCCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 1419
DB 61 GTTACTGCTCCTCCAGGTGGTACCGATACGTGATTATCAGAGAGCCACCAACCACTACT 120
QY 1420 GTCACTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTGTACTGCTCCT 1479
DB 121 GTCACTACTACTGAATATTGGTACAAATCCTTTGCTACTACTACTACTGTGTACTGCTCCT 180
QY 1480 CCAGTGGTACTGACTCAGTAATTTATCAGAGAACCCCAATCACTGTCACCTACCAACC 1539
DB 181 CCAGTGGTACTGACTCAGTAATTTATCAGAGAACCCCAATCACTGTCACCTACCAACC 240

QY 1540 GAGTATTGGTCTCAATCCTTTCTACTACTACTAGTACTGCTCTCCAGGTGGTACT 1599
DB 241 GAGTATTGGTCTCAATCCTTTCTACTACTACTAGTACTGCTCTCCAGGTGGTACT 300
QY 1600 GACTCAGTAATTTATCAGAGAACCTCCAAACCAACTGTCCACCACTGAATTTGGTCC 1659
DB 301 GACTCAGTAATTTATCAGAGAACCTCCAAACCAACTGTCCACCACTGAATTTGGTCC 360
QY 1660 CAATCTTAGCGAACCACTACTGCTGCTCTCTCCAGGAGGCACTGACTCAATTT 1719
DB 361 CAATCTTAGCGAACCACTACTGCTGCTCTCTCCAGGAGGCACTGACTCAATTT 420
QY 1720 ATCAGAGAACCACTACTGCTGCTCTCTCTCCAGGAGGCACTGACTCAATTT 1779
DB 421 ATCAGAGAACCACTACTGCTGCTCTCTCTCCAGGAGGCACTGACTCAATTT 480
QY 1780 ACCACTACACTGTAACTGACCACTGCTGCTCTCTCCAGGAGGCACTGACTCAATTT 1839
DB 481 ACCACTACACTGTAACTGACCACTGCTGCTCTCTCCAGGAGGCACTGACTCAATTT 540
QY 1840 CCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1899
DB 541 CCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 1900 GTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1959
DB 601 GTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 1960 GTCACACTACTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2019
DB 661 GTCACACTACTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 2020 CTTGTTGAACCTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2079
DB 721 CTTGTTGAACCTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 2080 GAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2139
DB 781 GAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 2140 GATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2199
DB 841 GATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 2200 CAATCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2259
DB 901 CAATCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 2260 ATTAGAGGCCACCAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2319
DB 961 ATTAGAGGCCACCAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 2320 ACACCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2370
DB 1021 ACACCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071

RESULT 3
US-08-357-962-5
; Sequence 5, Application US/08357962
; Patent No. 568263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA

COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
FILING DATE: 16-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICANT NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jervils, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-357-962-5

Query Match 9.2%; Score 348.2; DB 1; Length 425;
Best Local Similarity 92.8%; Pred. No. 5.5e-76;
Matches 376; Conservative 0; Mismatches 28; Indels 1; Gaps 1;
QY 1980 GTCTCAATCATATGCAACCACTACTACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2039
DB 21 GTCACATCATATGCAACCACTACTACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 80
QY 2040 TCTT-ATCAGAGAGCCACCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2098
DB 81 TTATCATTTAGAGAGCCACCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140
QY 2099 ATGCTACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2158
DB 141 ATGCAACCACTACTACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
QY 2159 AGCCACCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2218
DB 201 AGCCACCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 260
QY 2219 CCAGTCTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2278
DB 261 CCAGTCTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 320
QY 2279 CAACAGTCTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2338
DB 321 CAACAGTCTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 380
QY 2339 CTCTCCAGTGGTACTGACACTGCTGATTTATCTATGAAAGCATGT 2383
DB 381 CTCTCCAGTGGTACTGACACTGCTGATTTATCTATGAAAGCATGT 425

RESULT 4
US-08-878-106-5
; Sequence 5, Application US/08878106
; Patent No. 5817466
; GENERAL INFORMATION:

STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,962
FILING DATE: 16-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-357-962-2

Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;
QY 1300 CCAATCCAACTGTTAGTACTACTGATATTTGGTCTCAGTCTTCTGCTACACCACTACA 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 CCAATCMHACMGTYASYACHACYGARTAYTGTCNCARTCTNTWYCHACHACHACHACH 60
QY 1360 GTTACTGCTCTCCAGGTGGTACCGGATCTGCTGATTCAGAGGCA 1407
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RTDACYGSCWCCWGRGHACYGAYWCHGTDMTYATYAGRGCCW 108

RESULT 12
US-08-878-106-2
Sequence 2, Application US/08878106
Patent No. 5817466
GENERAL INFORMATION:
APPLICANT: Hoyer, Lois
APPLICANT: Livi, George
APPLICANT: Shatzman, Allan
TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,106
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/357,962
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: P50278
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5019
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-878-106-2
Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;
QY 1300 CCAATCCAACTGTTAGTACTACTGATATTTGGTCTCAGTCTTCTGCTACACCACTACA 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 CCAATCMHACMGTYASYACHACYGARTAYTGTCNCARTCTNTWYCHACHACHACHACHACH 60
QY 1360 GTTACTGCTCTCCAGGTGGTACCGGATCTGCTGATTCAGAGGCA 1407
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RTDACYGSCWCCWGRGHACYGAYWCHGTDMTYATYAGRGCCW 108
RESULT 13
US-08-145-705A-32/C
Sequence 3, Application US/08145705A
Patent No. 5489513
GENERAL INFORMATION:
APPLICANT: Springer, Wolfgang; Plempel, Manfred;
APPLICANT: Lberding, Antonius
TITLE OF INVENTION: SPECIFIC GENE PROBES AND
TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
TITLE OF INVENTION: INVESTIGATION OF CANDIDA
TITLE OF INVENTION: ALBICANS
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
COMPUTER: NEC PowerMate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/145,705A
FILING DATE: October 28, 1993
CLASSIFICATION: 535
PRIOR APPLICATION DATA:
APPLICATION NUMBER: German P 42 36 708.5
FILING DATE: October 30, 1992
ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Candida albicans

US-08-145-705A-32

Query Match 2.2%; Score 82.4; DB 1; Length 100;

Best Local Similarity 89.0%; Pred. No. 2.5e-11;

Matches 89; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1170 TGTATTGTTGCGATATCATCTACACCAACTGTTACCAAGTGAATGGACGGAAC 1229

Db 100 TGTATTGTTGCGATATCATCTACACCAACTGTTACCAAGTGAATGGACGGAAC 41

Qy 1230 AATCACTACCACTGCTACCACTGTTACCAAGTGAATGGACGGAAC 1269

Db 40 AATCACTACCACTGCTACCACTGTTACCAAGTGAATGGACGGAAC 1

RESULT 14

US-08-145-705A-34/c

Sequence 34, Application US/08145705A

Patent No. 5489513

GENERAL INFORMATION:

APPLICANT: Springer, Wolfgang; Plempel, Manfred;

APPLICANT: L bberding, Antonius

TITLE OF INVENTION: SPECIFIC GENE PROBES AND

TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC

TITLE OF INVENTION: INVESTIGATION OF CANDIDA

TITLE OF INVENTION: ALBICANS

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPRUNG HORN KRAMER & WOODS

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage

COMPUTER: NEC PowerMate 1 Plus

OPERATING SYSTEM: DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/145,705A

FILING DATE: October 28, 1993

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: German P 42 36 708.5

FILING DATE: October 30, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 8885-KGB

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Candida albicans

US-08-145-705A-34

Query Match 2.2%; Score 82.4; DB 1; Length 100;

Best Local Similarity 89.0%; Pred. No. 2.5e-11;

Matches 89; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 970 GGTATTGTCATTGTTGCTACCAACTAGACAGTTACAGACAGTACCACTGCTGTCTACTACT 1029

Db 100 GGTATTGTCATTGTTGCTACCAACTAGACAGTTACAGACAGTACCACTGCTGTCTACTACT 41

Qy 1030 TTACCACTCAATCCCAAGTGTGATAAAACCAACCAATCG 1069

Db 40 TTACCACTCAATCCCAAGTGTGATAAAACCAACCAATCG 1

RESULT 15

US-08-145-705A-33/c

Sequence 33, Application US/08145705A

Patent No. 5489513

GENERAL INFORMATION:

APPLICANT: Springer, Wolfgang; Plempel, Manfred;

APPLICANT: L bberding, Antonius

TITLE OF INVENTION: SPECIFIC GENE PROBES AND

TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC

TITLE OF INVENTION: INVESTIGATION OF CANDIDA

TITLE OF INVENTION: ALBICANS

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPRUNG HORN KRAMER & WOODS

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage

COMPUTER: NEC PowerMate 1 Plus

OPERATING SYSTEM: DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/145,705A

FILING DATE: October 28, 1993

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: German P 42 36 708.5

FILING DATE: October 30, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 8885-KGB

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

```
; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
US-08-145-705A-33

Query Match      2.0%  Score 76; DB 1; Length 100;
Best Local Similarity 85.0%  Pred. No. 9.3e-10;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 1070 AAATTTTGCAACCTATTCCAACCACTACCATCACAACCTTCATATGTTGGTGTGACTACTT 1129
      ||||||||||| ||||||| ||||||||||| ||||||||||| ||||||||||| |||
Db 100 AAATTTTGCAACCTATTCCAACCACTACCATCACAACCTTCATATGTTGGTGTGACTACTT 1129
      ||||||||||| ||||||| ||||||||||| ||||||||||| ||||||||||| |||

Oy 1130 CCTATCTGACTTAAGACTGCACCAATTTGGTGAACAGCTAC 1169
      ||||  || |||| |||| || |||| |||| |||| ||||
Db 40 CCTACAGAACCACCAACTGTACCAATAGGACAAACTGCTAC 1
```

Search completed: April 29, 2003, 19:20:50
Job time : 159 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2003, 19:18:29 ; Search time 259 Seconds
(without alignments)
15906.005 Million.cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgcttcacacattacatt.....tgccttggtcatttagtga 3786

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	78.4	2.1	5403	10 US-09-745-008-33	Sequence 33, Appl
C 2	74.4	2.0	684973	10 US-09-263-959-1	Sequence 1, Appl
C 3	62.2	1.6	2014	10 US-09-842-552-22	Sequence 22, Appl
C 4	55.4	1.5	436	10 US-09-864-761-2885	Sequence 2885, Ap
C 5	54.6	1.4	867	10 US-09-216-393-340	Sequence 340, App
C 6	54.6	1.4	867	10 US-09-216-393-342	Sequence 342, App
C 7	54.6	1.4	1397	10 US-09-216-393-343	Sequence 343, App
C 8	54.6	1.4	1397	10 US-09-216-393-345	Sequence 345, App
C 9	52.8	1.4	4104	10 US-09-801-368-107	Sequence 107, App
C 10	52.2	1.4	1059	12 US-10-073-256-54	Sequence 54, Appl
C 11	52	1.4	2015	10 US-09-842-552-79	Sequence 79, Appl
C 12	50.8	1.3	15720	9 US-10-025-380-1058	Sequence 1058, Ap
C 13	50.8	1.3	15720	10 US-09-922-217-1058	Sequence 1058, Ap
C 14	50.8	1.3	15720	10 US-09-833-263-1058	Sequence 1058, Ap
C 15	50.6	1.3	1236	9 US-10-077-584-3	Sequence 3, Appli
C 16	50.2	1.3	4197	10 US-09-137-531-7	Sequence 7, Appli
C 17	50.2	1.3	4197	10 US-09-137-531-8	Sequence 8, Appli
C 18	48.8	1.3	3183	10 US-09-955-909-1	Sequence 1, Appli
C 19	48.6	1.3	2614	9 US-09-822-846-491	Sequence 491, App

C 20	48.2	1.3	1101	10 US-09-874-062-2	Sequence 2, Appli
C 21	48.2	1.3	6604	10 US-09-880-107-1748	Sequence 1748, Ap
C 22	48	1.3	3331	10 US-09-864-761-19481	Sequence 19481, A
C 23	48	1.3	7104	10 US-09-815-242-4580	Sequence 4580, Ap
C 24	48	1.3	7107	10 US-09-815-242-8291	Sequence 8291, Ap
C 25	47.8	1.3	460	10 US-09-864-761-19383	Sequence 19383, A
C 26	47.8	1.3	2824	12 US-10-124-557-13	Sequence 13, Appl
C 27	47.8	1.3	3066	12 US-10-124-557-83	Sequence 83, Appl
C 28	47.8	1.3	3117	12 US-10-124-557-73	Sequence 73, Appl
C 29	47.8	1.3	3148	12 US-10-124-557-57	Sequence 57, Appl
C 30	47.8	1.3	3420	12 US-10-124-557-103	Sequence 103, App
C 31	47.8	1.3	3813	12 US-10-124-557-43	Sequence 43, Appl
C 32	47.8	1.3	3936	12 US-10-124-557-41	Sequence 41, Appl
C 33	47.8	1.3	3942	12 US-10-124-557-141	Sequence 141, App
C 34	47.8	1.3	3945	12 US-10-124-557-49	Sequence 49, Appl
C 35	47.8	1.3	3963	12 US-10-124-557-45	Sequence 45, Appl
C 36	47.8	1.3	3963	12 US-10-124-557-59	Sequence 59, Appl
C 37	47.8	1.3	4055	12 US-10-124-557-47	Sequence 47, Appl
C 38	47.8	1.3	4086	12 US-10-124-557-39	Sequence 39, Appl
C 39	47.8	1.3	4092	12 US-10-124-557-51	Sequence 51, Appl
C 40	47.8	1.3	4215	12 US-10-124-557-61	Sequence 61, Appl
C 41	47.8	1.3	4575	12 US-10-044-090-303	Sequence 303, App
C 42	47.8	1.3	5008	12 US-10-124-557-1	Sequence 1, Appli
C 43	47	1.2	5361	9 US-09-742-096-2	Sequence 2, Appli
C 44	47	1.2	6060	7 US-08-781-986A-534	Sequence 534, App
C 45	47	1.2	6152	9 US-09-742-096-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-745-008-33/c
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruci-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 5403
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33

Query Match	2.1%	Score 78.4;	DB 10;	Length 5403;
Best Local Similarity	42.2%	Pred. No. 4.2e-07;		
Matches 442;	Conservative 0;	Mismatches 606;	Indels 0;	Gaps 0;
QY 1309	ACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACACCACTACAGTTACTGCT	1368		
DB 3789	ACCGTGGGCACACTGCTGTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACACTGCTGTCAGCGGG	3730		
QY 1369	CCCTCAGGTGGTACCGGATACCTGTGATTATCAGAGAGACCAACCACTACTGTCTACTACT	1428		
DB 3729	AGTCGAGGGCGTACCGTGGGCACACTGCTGTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCAC	3670		
QY 1429	ACTGAATATTGGTGCACAACTCTTTGCTACTACTACTACTGTACTGTCTCTCCAGGTGGT	1488		
DB 3669	GCTGCAACGGGAGTGTGTCAGCGGGCGTACTGTGGGCACACTGCTGTGTCAGCGGGAGTCGAGGGCG	3610		
QY 1489	ACTGCACTCAGTATTATCAGAGAACCAACCACTCACTGTCTACTACACCACTAGTATTGG	1548		
DB 3609	ACCGTGGGCACACTGCTGTGTCAGCGGGAGTTGAGGGCGTACTGTGGGCACACTGCTGTCAACGGG	3550		

QY 1409 CAAACCATCTACTACTACTGAATATTTGGTCACAAATCCTTTCTACTACTACTACTG 1468
Db 13225 CTACAAATACCTGATGCTACACACTAGTAATTAATGTTCTTCCATAACCACTAGCTTCTC 13284
QY 1469 TTACTGCTCCTCCAGGTGGTACTGACTCAGTAATATATCAGAGAACCAACCAATCCAATG 1528
Db 13285 CTACAAGTAGTACTACTGTGAGTACTATTTGGTACCGTTCCCAATTTTCAGTGACTCCTTCTC 13344
QY 1529 TCATACACACCGAGTATTTGGTCTCAATCCTTTGGTCTACTACTACTACTGACTGCTCCTC 1588
Db 13345 TGACAAGTACTGCTGATGCCACCAATAGTACTACTGACTATTTATTTGGCCACTACTTCTTCTC 13404
QY 1589 CAGGTGGTACTGACTCAGTAATATCAGAGAACCTCCAAACCAACTGTCACCAACAGT 1648
Db 13405 TAACAGGACTACTGATGTTAGCACTACTACTACTATTAATAATAAGTACTCCTGTTTC 13464
QY 1649 ATATTGGTCCCAATCTTACGCAACCACTACTGACTGCTCCTCCAGGAGCACTG 1708
Db 13465 AAACAATACTACTAATGCTAGCACTAGTACTAATTTGGCTAATAATTAATGCTACTCTTC 13524
QY 1709 ACTCAGTAATATCAGAGAACCAACCAACCACTGTCACACTACTACTGTAATACTGTCAC 1768
Db 13525 ATACAAGTACTGATGATGTTCTTAATAATGTTCCAGTACTAGCTATTTCTTCTC 13584
QY 1769 AATCATATGCCACCACTACTGTAATGTCACCAAGGTGGTACTGACACTGTTATC - 1827
Db 13585 TTGCAAACTACTGTTGACACTACTAGCAACAGTTTTTCCATTAATGACCACTTCTTCT 13644
QY 1828 -----ATTAGAGAGCCCAACCACTACTGTCACACTACTGATGATTTGGTCTCAATCGT 1882
Db 13645 CTGAAAGTACTAATGCTATGAACACTACTGTTATTATGGCAACTACTTCTCCTACAACTA 13704
QY 1883 TTGCTACTACCACTGTAATGCTGTCACCAAGTGGCACTGATGCTGTTATCAATAGG 1942
Db 13705 CTGATGTTGCTAGCACAAATATGATGTTCTATGACAAATTTCTTTAGCTACAATG 13764
QY 1943 AACCAACCAACCACTGTCACCACTACTGTAATGTCACCAAGTGGTCTCAATCATATGCAACCACTA 2002
Db 13765 CTGCTGGTAATAATACTAGTAATGTTATTCATTAACCACTACTTCTTTGGTAATAGT 13824
QY 2003 CTACATACCGCTCCACTGTTGAACTGATACCGTTCTTATCAGAGAGCCCAACCACT 2062
Db 13825 TTCTTTGTTGACTACTCTCTCCAACTGATGCTGATGCTACTACTACAAGTAATACTA 13884
QY 2063 ATACTGCTACTACTGTAATGCTGTCATATATGCTACAACCACTGTTACTG 2122
Db 13885 ATCTTGGCATGACTACTATTATGACAGCTTCTCTAGCAATTCCTACCCATACTTACTT 13944
QY 2123 CACCACCTGGTGAACCGATACCGTTCTATCAGAGAGCCCAACCACTACTGTCAT 2182
Db 13945 CTATTCTAGCTCTATTACTTCTATTTTGGATGATGTTTCCAACTAATACTACTA 14004
QY 2183 CTACTGAA 2190
Db 14005 CTGATAAA 14012

RESULT 3
US-09-842-552-22/c
; Sequence 22, Application US/09842552
; Patent No. US20020055628A1
; GENERAL INFORMATION:
; APPLICANT: The Regents of The University of California
; TITLE OF INVENTION: MULTILOCUS REPETITIVE DNA SEQUENCES FOR GENOTYPING BACILLUS ANTHR
; TITLE OF INVENTION: RELATED BACTERIA
; FILE REFERENCE: S-89,687
; CURRENT APPLICATION NUMBER: US/09/842,552
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/199,911
; PRIOR FILING DATE: 2000-04-26
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22

; LENGTH: 2014
; TYPE: DNA
; ORGANISM: Bacillus anthracis
US-09-842-552-22
Query Match 1.6%; Score 62.2; DB 10; Length 2014;
Best Local Similarity 42.3%; Pred. No. 0.0093;
Matches 402; Conservative 0; Mismatches 548; Indels 1; Gaps 1;
QY 1354 ACTACAGTTACTGCTCCTCCAGGTGGTACCGATACCTGCTATATATCAGAGAGCCACCAAC 1413
Db 1356 ACAACTGGTGTGTTCTCTACGAGTGGTCTCTCTGCAACTGGTGTGTTCTTCTAGCATC 1297
QY 1414 CATACTGTCACACTACTGTAATTTGGTGCACAAATCCTTTGCTACTACTACTACTGTTACT 1473
Db 1296 GATGTTTCTCTGCAACTGGTGTGTTCTCTACGACCGATGTTTCCCTCTGCAACTGGTGT 1237
QY 1474 GCTCTCCAGGTGGTACTGACTCAGTAATATCAGAGAACCAACCAATCAACTGTCAT 1533
Db 1236 TCTTCTGCAATCGGTCTTCTTCTACAACTGATTGCTCTTACGACCGGTGCTTCTTCT 1177
QY 1534 ACAACCGAGTATTTGGTCTCAATCCTTTGCTACTACTACTACTACTGCTCTCTCCAGGT 1593
Db 1176 ACAATCGGTGTTCTTCTACAACTGATGTTCTTCTACGACCGGTGCTTCTTCTGCACT 1117
QY 1594 GGTACTGACTCAGTAATATATCAGAGAACCTCCAAACCCAACTGTCACCACTGTAATAT 1653
Db 1116 GGTGTTCTTCTACAACTCGGTGTTCTTCTACAACTGATTGTTCTTCTAGCACCGGTGCT 1057
QY 1654 TGGTCCCAATCTTACGCAACCACTACTGTCAGTCTGCTGCTCCAGGAGGCACTGACTCA 1713
Db 1056 TCTTCTGCAACTGGTGTGTTCTTCTACAACTGGTGTGTTCTTCTTCTACAACTGATTGTTCTC 997
QY 1714 GTAATATCAGAGAACCAACCAACCACTGTCACACTACTACTGTAATACTGGTCAACATC 1772
Db 996 GCAATTAACCACTCGCAAGCTTTGAACTATCTTTTCTGTTGTTCTATCAGCACGCA 937
QY 1773 ATATGCCACCACTACCACTGTAATGTCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1832
Db 936 ACATGATGAATTCATTTCTTCTCAATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 877
QY 1833 AGAGCCCAACCACTGTCACACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1892
Db 876 CTCAAGTGTACAGGTTCACACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
QY 1893 CACAAGTGTAACTGGTCCCAAGTGGCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1952
Db 816 TTCTGACTCTTCTGTTCTTCTGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 757
QY 1953 CCAACTGTCACCACTACTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2012
Db 756 TTCTGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 697
QY 2013 CGCTCCACCTGGTGAACCTGATACCGTCTTATCAGAGAGCCCAACCACTGCTGCTGCTGCT 2072
Db 696 TAATCTTCTGTTCTGCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 637
QY 2073 TACTACTGTAATGCTGCTCAATCATATGCTACAACCACTGTTACTGCTGCTGCTGCTGCTGCT 2132
Db 636 TTCTACTTCTTCTGACTCTTCTGTTCTGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCT 577
QY 2133 TGAACCGGATACCGTGTGTTATCAGAGAGCCCAACCACTGCTGCTGCTGCTGCTGCTGCTGCT 2192
Db 576 TTCCGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 517
QY 2193 CTGGTCTCAATCATGCTACACCACTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2252
Db 516 TGACTCTTCTGTTCTGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 457
QY 2253 TGTATCATTAAGAGAGCCCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2303
Db 456 TTCTACTTCTTCTGACTCTTCTGTTCTTCTGCAATTAATTTCTTCTTCTTCTTCTTCTTCTTCT 406

Sequence 79, Application US/09842552
Patent No. US20020055628A1
GENERAL INFORMATION:
APPLICANT: The Regents of The University of California
TITLE OF INVENTION: MULTICUS REPTITIVE DNA SEQUENCES FOR GENOTYPING BACILLUS A.
FILE REFERENCE: S-89,687
CURRENT APPLICATION NUMBER: US/09/842,552
CURRENT FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/199,911
PRIOR FILING DATE: 2000-04-26
NUMBER OF SEQ ID NOS: 106
SOFTWARE: Patent in version 3.0
SEQ ID NO 79
LENGTH: 2015
TYPE: DNA
ORGANISM: Bacillus anthracis
US-09-842-552-79

Query Match 1.4%; Score 52.2; DB 12; Length 1059;
Best Local Similarity 41.6%; Pred. No. 0.11;
Matches 333; Conservative 0; Mismatches 468; Indels 0; Gaps 0;

QY 1303 AATCCAACTGTTAGTACTACTGAATATGCTCTCAGTCCCTTGTCTACAAACCACTACAGTT 1362
DB 1010 ACTTACCTTTAGGATTTGGTGGTCTTCTGTTTATTTGGTGGTGAACCTTGAA 951
QY 1363 ACTGCTCCCTCCAGGTGATACCGATCTGTTGATATCAGAGAGCCACCAACCACTACTGTC 1422
DB 950 TCGGAGTCACTATCTGAGTCTGAGTGGCTATCTGAATCCGAGTCGCTATCCGAGTCTGAG 891
QY 1423 ACTACTACTGAATATTTGGTGCACAACTCTTTGCTACTACTACTGTTACTGCTCTCCCA 1482
DB 890 TCGCTATCTGAGTCTGAATCGCTGCTGAGTCTGAGTCGCTATCCGAGTCTGAGTCGCTG 831
QY 1483 GGTGGTACTGACTCAGTCAATATATCAGAGAACCAACCAATCCAACTGTCTACAAACCGAG 1542
DB 830 TCTGAATCTGAATCACTGCTGAATCCGAATCGCTATCTGAATCTGAATCGCTATCCGAG 771
QY 1543 TATTGGTCTCAATCCCTTTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 1602
DB 770 TCTGAGTCTGAGTCTGAATCGCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 711
QY 1603 TCAGTAATATCAGAGAACCTCAACCCCACTGTCACCACTGTCACCACTGTCACCACTGTC 1662
DB 710 TCGCTGCTGAGTCTGAATCGCTATCTGAATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 651
QY 1663 TCTTACGCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1722
DB 650 TCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 591
QY 1723 AGAGAACCAACCACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1782
DB 590 TCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 531
QY 1783 ACTTACCACTGTAATCTGACCACTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1842
DB 530 TCGCTATCTGAGTCTGAATCGCTATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 471
QY 1843 AACCACACTGCTACTACTGAGTATTTGCTCAATCTGCTCAATCTGCTCAATCTGCTCAAT 1902
DB 470 TCTGAATCTGACTCACTATCTGATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 411
QY 1903 ACTGGTCCCAAGTGGCACTGATCTGTTATCATTTAGGAGAACCAACCAACCACTGTC 1962
DB 410 TCTGAATCACTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGA 351
QY 1963 ACCACTACTGATCTGCTCAATCATATGCAACCACTACTACTACTACTACTACTACTACT 2022
DB 350 TCGCTATCTGAATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 291
QY 2023 GGTGAATCTGATACCGTCTTATCAGAGAGCCCAACCAACCACTACTGTCATCTACTGAA 2082
DB 290 TCTGAATCTGAATCACTGCTGAATCCGAATCGCTATCTGATCTGATCTGATCTGATCTGA 231
QY 2083 TACTGGTCTCAATCATATGCT 2103
DB 230 CTTGAGTCTGCTGAGCTGCTGAGCT 210

RESULT 11
US-09-842-552-79/c

Sequence 79, Application US/09842552
Patent No. US20020055628A1
GENERAL INFORMATION:
APPLICANT: The Regents of The University of California
TITLE OF INVENTION: MULTICUS REPTITIVE DNA SEQUENCES FOR GENOTYPING BACILLUS A.
FILE REFERENCE: S-89,687
CURRENT APPLICATION NUMBER: US/09/842,552
CURRENT FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/199,911
PRIOR FILING DATE: 2000-04-26
NUMBER OF SEQ ID NOS: 106
SOFTWARE: Patent in version 3.0
SEQ ID NO 79
LENGTH: 2015
TYPE: DNA
ORGANISM: Bacillus anthracis
US-09-842-552-79

Query Match 1.4%; Score 52; DB 10; Length 2015;
Best Local Similarity 48.3%; Pred. No. 0.17;
Matches 145; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 1305 TCCAACTGTTAGTACTACTGAATATGCTCTCAGTCTCTGCTTGTGCTACAAACCACTACAGTTAC 1364
DB 1370 TTCTTCTTTTGTACAACTGGTGTCTTCTTACGACTGGTCTCTCTGCAACTGGTGTG 1311
QY 1365 TGTCTCTCCAGGTGGTACCGATCTGATATCAGAGAGCCACCAACCACTACTGTCAC 1424
DB 1310 TTCTTCTTACGATCGATGTTTCTCTGCAACTGGTGTGTTCTTCTACGACCGATGTTTCTCTC 1251
QY 1425 TACTACTGAATATGTTGCTCAACATCTTTGCTACTACTACTACTACTACTACTACTACTACT 1484
DB 1250 TGAACCTGGTGTGTTCTTCTGCAATCGGTGCTTCTTCTTACAACTGATGTTGTTCTTCTAC 1191
QY 1485 TGTACTGACTCAGTATATCAGAGAACCAACCACTGCTCACTACTACTACTACTACTACTACT 1544
DB 1190 CGGTGCTTCTTCTTACAACTGGTGTGTTCTTCTTACAACTGATGTTGTTCTTCTACGACCG 1131
QY 1545 TTGCTCTCAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1604
DB 1130 TTCTTCTGAGTGGTGTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1071

RESULT 12
US-10-025-380-1058
Sequence 1058, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1058

; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1058
; LENGTH: 15720
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-833-263-1058

Query Match 1.3%; Score 50.8; DB 10; Length 15720;
Best Local Similarity 43.8%; Pred. No. 0.98;
Matches 269; Conservative 0; Mismatches 342; Indels 3; Gaps 1;
QY 1717 ATTATCAGAGAACCAACCAACCACTGTCACCTACTACTGTAATACCTGGTCCACAAATCATAT 1776
DB 4756 ACTCCAGCCCTCCCAACACCACTCCAGTCCTCCAAACATCACCACCAACCCCT 4815
QY 1777 GCCACACTACCTACTGTAAGCAGCAGCAGGCTGAGACTGTTATCATTTAGAGAG 1836
DB 4816 CCACCAACCACTCCAGCCCTCCCAACCAACGACCAACCAACCCCTCCACCAACCC 4875
QY 1837 CCACCAACCACTGTCACCTACTACTGATGATTGCTCAATCGTTTGTCTACTACCA 1896
DB 4876 ACTCCAGCC--CTCAACGACTACACCCATCCTCCCAACCAACCAACCCCTCCAC 4932
QY 1897 ACTGTAACTGGTCCACCAAGTGGCACTGATCTGTTATCATTAAGGGAACCAACCA 1956
DB 4933 CCACCAACCACTCCAGCCCTCCACCAACCAACCAACCAACCCCTCCACCAACCC 4992
QY 1957 ACTGTACCACTACTGAATACCTGCTCAATCATATGCAACCACTACTACCACTACCG 2016
DB 4993 ACTCCAGCCCTCCCAACCAACCACTCCAGCCCTCCCAATACCAACCAACCAACCC 5052
QY 2017 CCACCTGGTGAATCATGATACCGTCTTATCAGAGAGCCCAACCAACCACTACTG 2076
DB 5053 CCACCAACCACTCCAGCTCTCAATACCAACCACTCCAGCCCTCCCAACCAACCA 5112
QY 2077 ACTGAATACCTGCTCAATCATATGTCACCAACCACTGTTACTGCACCACTGGTGA 2136
DB 5113 ATGACCAACCCCTCCACCAACCACTCCAGCTCTCCCAATACCAACCAACCAACCC 5172
QY 2137 ACCGATACCTGCTTATCAGAGAGCAGCAACCACTACTGCTACTACTGATGATGG 2196
DB 5173 TCCTCACTACCACTCCAGCCCTCCCAACCAACCACTGACCAACCACTCCACCAAC 5232
QY 2197 TCCTCAATCATATGTCACCAACCACTGTTACTGCACCACTGGTGGTACCGATCTGT 2256
DB 5233 ACTCCAGCCCTCCACCAACCACTGACCAACCCCTTCCACCAACCACTCCAGCCCT 5292
QY 2257 ATCATTAGAGAGCCCAATCAACAGTTACTACTACTGTAATTTGGTCAACAACTTT 2316
DB 5293 CTAAACAATCTCTCTACCTCCATCAATAACTCTCTCTTACATTTTCCACCACTTCA 5352
QY 2317 GCCACCAACCACTAC 2330
DB 5353 ACAACCCCTACTAC 5366

RESULT 15
US-10-077-584-3
; Sequence 3, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
; APPLICANT: LINDQUIST, SUSAN
; APPLICANT: KROBITSCH, SYLVIA
; TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
; FILE REFERENCE: ARCD:367US
; CURRENT APPLICATION NUMBER: US/10/077,584
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269,157
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(513)
; US-10-077-584-3

Query Match 1.3%; Score 50.6; DB 9; Length 1236;
Best Local Similarity 48.1%; Pred. No. 0.26;
Matches 143; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 3220 GAATCAGCATCCCAAGTTTGAACCAAGTATGGGTGAAATTTCTGATTAACTACTTCT 3279
DB 57 GCAGCAACGCAACACAGCAGCAACCAACAGCAACAGCAACAGCAACAGCAACAGCA 116
QY 3280 ACTGAAATTTGAAGCTACAAACCAACCACTCCCTACAGAGCTCCATCACCTGCTGTTCT 3339
DB 117 ACAGCAACCAACAGCAGCAACCAACAGCAACAGCAACAGCAACAGCAACAGCA 176
QY 3340 GGTACTGATGTAATCTACTGAACTGATGATGATGATGATGATGATGATGATGATG 3399
DB 177 ACAACAGCAGCAACAGCAGCAACAGCAGCAACAGCAACAGCAACAGCAACAGCA 236
QY 3400 ACTTCAAAACCAACAGTGAACCTGGTGTGCTACTACAGCACTACTATGAAATGTTGT 3459
DB 237 GCAGCAACGCAACACAGCAGCAACCAACAGCAACAGCAACAGCAACAGCAACAGCA 296
QY 3460 AAATCTCCATCAACTGATTTAAACATCAAGCTTGACAAACAGCAGCACTCAGCATCTA 3516
DB 297 ACAGCAACCAACAGCAGCAACCAACAGCAACAGCAACAGCAACAGCAACAGCAAC 353

Search completed: April 29, 2003, 22:10:32
Job time : 569 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 29, 2003, 17:19:18 ; Search time 3337 Seconds
(without alignments)
18374.622 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgttcaacaattacatt.....tgtctgttcattagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	400.6	10.6	959	17 CNS07DAZ	AL440241 T3 end of
C 2	353	9.3	893	17 CNS07D66	AL440068 T7 end of
C 3	272.8	7.2	1050	17 CNS07CMQ	AL439368 T7 end of
C 4	164.6	4.3	1047	17 CNS07D6L	AL440083 T7 end of
C 5	144.4	3.8	1011	17 CNS07DAY	AL440240 T7 end of
C 6	102	2.7	735	17 CNS04NSM	AL299119 Tetraodon

7	100.4	2.7	773	17	CNS01VTG	AL169549 Tetraodon
8	98.4	2.6	908	17	CNS07DJI	AL440548 T7 end of
9	97	2.6	762	17	CNS01XFI	AL171639 Tetraodon
10	94.8	2.5	824	17	AZ185454	AZ185454 SP_1005A
11	91.4	2.4	501	17	FR0048173	AL444958 Fugu rubr
12	91.2	2.4	989	17	CNS02HA4	AL197365 Tetraodon
13	89.6	2.4	450	17	FR0048073	AL444858 Fugu rubr
14	82.6	2.2	424	17	FR0025683	AL018519 F.rubripe
15	81.4	2.2	641	13	BM181884	BM181884 fv51b11.y
16	76.2	2.0	694	17	BH355163	BH355163 CH230-81P
17	75.6	2.0	530	17	AZ166409	AZ166409 SP_0088A
18	75.6	2.0	619	17	FR0047601	AL444386 Fugu rubr
19	75	2.0	500	17	B67199	B67199 Cpe0015B Cp
20	74	2.0	935	17	CNS033D4	AL225985 Tetraodon
21	73.4	1.9	619	17	FR0006944	Z90754 F.rubripes
22	72.6	1.9	644	13	BM605179	BM605179 170006870
23	72	1.9	485	17	AZ34703	AZ34703 1M0221C24
24	70.4	1.9	521	17	AZ164800	AZ164800 SP_0076B
25	69.6	1.8	350	17	BH879665	BH879665 ht47h12.g
26	69.6	1.8	740	10	AV973760	AV973760 AV973760
27	69.2	1.8	788	17	AZ183942	AZ183942 SP_1002A
28	68.8	1.8	331	17	BH881398	BH881398 hv25c11.b
29	68.4	1.8	933	17	AZ204694	AZ204694 SP_0100A
30	68	1.8	470	17	FR0018463	AL011359 F.rubripe
31	67	1.8	810	17	AZ193472	AZ193472 SP_1039B
32	66.4	1.8	358	17	BH777606	BH777606 fzmb013f0
33	65.6	1.7	605	17	AZ640388	AZ640388 1M0502E21
34	65	1.7	642	17	CI1688	AJ226509 Clona Int
35	64.2	1.7	778	13	BM170064	BM170064 EST572587
36	64	1.7	914	17	AZ203202	AZ203202 SP_0100A
37	63.4	1.7	695	13	BJ388152	BJ388152 BJ388152
38	62.4	1.6	273	17	FR0047466	AL444251 Fugu rubr
39	62.4	1.6	619	17	FR0013713	AL004959 F.rubripe
40	62	1.6	415	10	BE56771	BE56771 fk95c07.y
41	61.6	1.6	979	17	CNS06PDW	AL409242 T7 end of
42	61.4	1.6	550	17	FR0043207	AL130899 Fugu rubr
43	61.4	1.6	718	17	AZ972907	AZ972907 2M0246M24
44	61.2	1.6	619	17	FR0006038	Z89848 F.rubripes
45	61.2	1.6	832	13	BM170146	BM170146 EST572669

ALIGNMENTS

RESULT 1	CNS07DAZ/c	CNS07DAZ	959 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T3 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of					
DEFINITION	Candida tropicalis, genomic survey sequence.					
ACCESSION	AL440241					
VERSION	AL440241.1	GI:12223652				
KEYWORDS	GSS.					
SOURCE	Candida tropicalis.					
ORGANISM	Candida tropicalis					
REFERENCE	1 (bases 1 to 959)					
AUTHORS	Souci, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies					
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)					
MEDLINE	20584711					
PUBMED	11152876					
REFERENCE	2 (bases 1 to 959)					
AUTHORS	Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.					
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis					

Db	75	GATTATGCAACTAGGTTTACTGAGACACATGTCAGACCAAACTGATGTTGTTGT 20
RESULT 4	CNS07D6L	1047 bp DNA linear GSS 08-JUL-2001
LOCUS	T7 end of clone BD0AA009H06 of library BD0AA from strain CBS 94 of	
DEFINITION	Candida tropicalis, genomic survey sequence.	
ACCESSION	AL440083	
VERSION	AL440083.1	GI:12223494
KEYWORDS	GSS.	
SOURCE	Candida tropicalis.	
ORGANISM	Candida tropicalis	
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
AUTHORS	Saccharomycetales; mitosporic Saccharomycetales; Candida.	
REFERENCE	1 (bases 1 to 1047)	
AUTHORS	Souci�t,J.L., Aigle,M., Artiguenave,F., Blandin,G.,, Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lior�nt�,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., ,Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.	
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies	
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)	
MEDLINE	20584711	
PUBMED	11152876	
REFERENCE	2 (bases 1 to 1047)	
AUTHORS	Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.	
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis	
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)	
MEDLINE	20584726	
PUBMED	11152891	
REFERENCE	3 (bases 1 to 1047)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqrefgenoscope.cns.fr - Web : This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.	
FEATURES	Location/Qualifiers	
Source	1..1047	
	/organism="Candida tropicalis"	
	/strain="CBS 94"	
	/db_xref="taxon:5482"	
	/clone="BD0AA009H06"	
	/clone.lib="BD0AA"	
	/note="end : 17"	
misc_feature	Complement(<2..>484)	
	/note="similar to O74660 [Agglutinin-like protein 4 precursor, ALS4] [Candida albicans]"	
	/evidence=not_experimental	
BASE COUNT	373 a 178 c 171 g 322 t	3 others
ORIGIN		
Query Match	4.3%;	Score 164.6; DB 17; Length 1047;
Best Local Similarity	60.5%;	Pred. NO. 2.7e-31;
Matches	291; Conservative	0; Mismatches 184; Indels 6; Gaps 1;
QY	2	TGCTTCAACAATTTACATTTGTTATTCTCTATATTTCATTTCCAACTGCAAGCAATCA 61
Db	483	TTCTATACAGCCATTGCTGTTAATTAATGATTCAATTTGATTTCTTCAAGAAGATAT 424

Db 75 GATTATCAACTAGGTTACTGAGACACATGGTCCAGACCAACTGATGTTGTTGT 20

RESULT 4	CNS07D6L	1047 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T7 end of clone BD0AA009H06 of library BD0AA	from strain CBS 94 of			
DEFINITION	Candida tropicalis, genomic survey sequence.				
ACCESSION	AL440083				
VERSION	AL440083.1	GI:12223494			
KEYWORDS	GSS.				
SOURCE	Candida tropicalis.				
ORGANISM	Candida tropicalis				
REFERENCE	1 (bases 1 to 1047)				
AUTHORS	Souciat,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boletini-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
PUBMED	11152876				
REFERENCE	2 (bases 1 to 1047)				
AUTHORS	Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis				
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)				
MEDLINE	20584726				
PUBMED	11152891				
REFERENCE	3 (bases 1 to 1047)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.				
FEATURES	Location/Qualifiers				
source	1..1047				
	/organism="Candida tropicalis"				
	/strain="CBS 94"				
	/db_xref="taxon:5482"				
	/clone="BD0AA009H06"				
	/clone_lib="BD0AA"				
	/note="end : T7"				
misc_feature	complement(<2..>484)				
	/note="similar to 074660 [Agglutinin-like protein 4 precursor, ALS4] [Candida albicans]"				
	/evidence-not_experimental				
BASE COUNT	373 a	178 c	171 g	322 t	3 others
ORIGIN					
Query Match	4.3%	Score 164.6;	DB 17;	Length 1047;	
Best Local Similarity	60.5%;	Pred. No. 2.7e-31;			
Matches 291;	Conservative 0;	Mismatches 184;	Indels 6;	Gaps 1;	
QY	2	TGCTTCAACCAATTTACATTTGTTTCTCTATATTTGTTCAATTTGCAAGTGCACCAAGCAATCA	61		
DB	483	TTCTCATACAGCCATTTGCTCTATTATATGATTCATTCATTCGTTATCTTCAAAAGATAT	424		

```
QY 62 CTGGTGTCTTTGATAGTTTAAATTCATTAACTTGGTCCCAATGCTGCTAATTA-----TG 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 CTGGGATTTTATGTTTCGATTTCATTAAACATGGAATGCTGCTCGGATTTACCCAGTG 364
QY 116 CTTTCAAGGGCCAGGATACCAACTTGAATGCTGTTTGGTGGTCCCTTAGATGGTA 175
Db 363 CATACCAAGTCCACAAATTTCTTACTGCGACGACAACTGAATGTTTGAATGGAG 304
QY 176 CCAGTGCCTCAATCCAGGGATACATTACATTTGAATGCGATGCTGTTTAAATATACTA 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 AATCTGCTGAACAGGTGATACATTACATTAATATGCGCATGTATTCAAGTTCATTA 244
QY 236 CTTCAAAATCATCTGTGTGATTTAACTGCGGATGGTGTAAATATGCTACTTGTCAATTTT 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 CAAATCAAAACACTGTGTGATTTGATGCTGATGGCACTACTTATGCGACCTGTAACTCA 184
QY 296 ATTCGTGTAAGAAATTCACAACTTTTCTACATTACATGCTACTGGAACGCGCTTTGA 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ATTCGTGGAAGAATTTTACTTCTCAAGTTTAAAGTGTACTGTTTCATCTACCTTGA 124
QY 356 AATCATCCATTAAGGCATTTGGTACAGTTTACTTTACCAATTCGATTCATGTTGGTGA 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 CACATATACCCAGCTGCTGGTACTTTACATGTTTCCCTTGACATTTAATGTTGGTAGTT 64
QY 416 CAGGTTTCATCAACTGATTTGGAAGATTTAAATGTTTACTGCTGTCACCAATACAGTCA 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 CGGGAACCTCTGTTGCTGACTGATTCCTACTGCTTCAGACAGGTGAACACAGTAA 4
QY 476 C 476
Db 3 C 3

RESULT 5
CNS07DAY 1011 bp DNA linear GSS 08-JUL-2001
LOCUS T7 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of
DEFINITION Candida tropicalis, genomic survey sequence.
ACCESSION AL440240
VERSION AL440240.1 GI:12223651
KEYWORDS GSS
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis.

REFERENCE
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1011)
AUTHORS Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 16. Candida
tropicalis
FEBS Lett. 487 (1), 91-94 (2000)
20584726
PUBMED 11152891
REFERENCE 3 (bases 1 to 1011)
AUTHORS Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqrefgenoscope.cns.fr - Web :
This GSS is part of a random genomic sequencing program of thirteen
```

```
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Location/Qualifiers
1..1011
/organism="Candida tropicalis"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone_lib="BD0AA010H12"
/clone_lib="BD0AA"
/notes="end : T7"
misc_feature
<52..>988
/notes="similar to P46590 [ Agglutinin-like protein 1
precursor, ALS1 ] [ Candida albicans ]
1 putative frameshift(s)"
/evidence="not_experimental"

BASE COUNT 336 a 188 c 151 g 293 t 43 others
ORIGIN

Query Match 3.8%; Score 144.4; DB 17; Length 1011;
Best Local Similarity 62.6%; Pred. No. 4.5e-26;
Matches 234; Conservative 4; Mismatches 135; Indels 1; Gaps 1;

QY 7 CAACAATTTACATGTTATTCCTATATTTGTCAATTTGCAAGTGCAAGACAATCAGTGT 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 CAAACAACCTCTCTATGCTGTGCATTTGATGCTACTTCAAGTAAAGAAATTTCCGGT 684
QY 67 GTTTTGTAGTATTTAAATTCATTAACCTTGCTCCAATGCTGCTAATATGCTTTCAAGGG 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 GTGTTTCATAGGATTTGAATCCCTTAACCTTGGGATAAGAGCTGGTAATATGATATCAAGGT 744
QY 127 CCAGGATACCCAACTTGGAAATGCTGTTTGGGTTGGTCCCTTAGATGGTACCAGTGCCAAT 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 CCTCAATATCCAAACATGGAATGCCGTGTTGATTTGGTTCGTAGATGSCACAACATCC 804
QY 187 CCAGGGGATACATTCACATGATATGCCATGTCGTTTAAATATACTACTTCAACAACA 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 CCAGGTGATACGTTACTTGTTCATGCTGCTTCAAGTTCATGCTAGTGTGCAAT 864
QY 247 TCTGTTGATTTAACTCCGATGCTGTTAAATATGCTACTTGTCAATTTTAT-TCTGTGA 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 CTGTTGGATGCTGATGCCAATGCKATGACATATGCCACATGTCATCTTCATGCKGGTGA 924
QY 306 AGAATTCACAACCTTTTCTACATTAACATGCTACTGTGAACGACGCTTTGAAATCATCCAT 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 AGAGTTTACTACTTACTCTAGMTTACGTTGCTGCTACTGCTAYAGATTCGTTAAGTCTGTTTCAT 984
QY 366 TAAGGCATTTGGTA 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 GAAGCTATGGGATA 998

RESULT 6
CNS04NSM/c 735 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 123M05 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL299119
VERSION AL299119.1 GI:8038260
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
```


Qy	1423	ACTACTACTGAATATTGGTCACAATCCCTTTGGCTACTACTACTACTGTTTACTGCTCCCTCCA	1482
Db	312	ACTACTGCTTCTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTTCTACT	371
Qy	1483	GGTGGTACTGACTAGTAATTATCAGAGAACCACCAATCCCAATGTCACCTACCAACCGAG	1542
Db	372	ACTGCTACTACTACTACTACTACTACTACTACTGCTACTACTACTACTAACTGCTACT	431
Qy	1543	TATTGGTCTCAATCCCTTTGCTACTACTACTACAGTTTACTGCTCCCTCAGGTGGTACT	1599
Db	432	ACTACTGCTGCTACTACTACTACTACTGCTAGTAGTACTACTACTACTACTAGTAGT	488

RESULT 12	CNS02HA4	989 bp	DNA	linear	GSS 13-MAY-2000
Locus	CNS02HA4				
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 139021 of library G from Tetraodon nigroviridis, genomic survey sequence.				

JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT	<p>characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis</p> <p>3 (bases 1 to 989)</p> <p>Unpublished Genoscope.</p> <p>Direct Submission</p> <p>Submitted (12-APR-2000)</p> <p>This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.</p>
--	--

```

/organism="retrovirus" /db_xref="taxon:9883" /clone="139021" /clone_lib="G" /note="Genoscope sequence ID : COAGI39AHillPl-end : T7" BASE COUNT      320 a    245 c   122 g    261 t     41 others ORIGIN          Query Match       2.4%; Score 91.2; DB 17; Length 989; Best Local Similarity 47.0%; pred.No. 2.8e-12; Matches 238; Conservative 9; Mismatches 259; Indels 0; Gaps 0: QY  3113 CTACAGGTGCATAATGGAGACAATACTTCATCAACCATCCAGTTCCTCCTGAAGCAAA 3172         ||||| ||||| ||||| ||||| ||||| ||||| Db  31 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 90 QY  3173 GTACTTTAGCATCTCGAAGTGAAAGAACAACAAAAGGGTTCTTCATGAATCAGCATCCA 3232         ||||| ||||| ||||| ||||| ||||| ||||| Db  91 CTACTACTACTACTACTACTACTTAATACAAAAAACACWACWACTACTACTACTACTA 150

```

```

Qy 3233 CAAGTTTGAACCAAGATGAGTGGTGAATTTCTGGATTAACCTACTTCTACTGAAATTTGAAG 3292
Db 151 CTAATACTACTACTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 210

Qy 3293 CTACAACAACAGTCCCTACAGAACCTCATACACCTGCTGTTTCTTCTGTTGATGATGAA 3352
Db 211 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTG 270

Qy 3353 CTACTGAACCAACTGATAGAACCAACCTACTACTACTACTACTACTACTACTACTACTACTACT 3412
Db 271 CTAATACTACTACTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 330

Qy 3413 ACAGTGAACCTGGTCTACTACTACAAGCTACTAATGAATGGTGGTAAATCTCCATCAA 3472
Db 331 CTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 390

Qy 3473 CTGATTTAAACATCAAGCTTGACACAGCAGCCTCAGCATCTACAGTGCTTAATAGCGAAC 3532
Db 391 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 450

Qy 3533 TTGTACTAGTGATCTGTTACTGTTGAGCTGTTGCCAGTCTTCAATGATCAATCAC 3592
Db 451 ATNNTNNTNGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 510

Qy 3593 ATTCTACTTCTGTTACCNACAGCAAC 3618
Db 511 ATAATAATAMTACTACTCCCTCCATC 536

RESULT 13
LOCUS FR0048073 494 bp DNA linear GSS 05-JAN-2001
DEFINITION Fugu rubripes GSS sequence, clone 263K15bD8, genomic survey
sequence.
ACCESSION AL444858
VERSION AL444858.1 GI:12052694
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 494)
AUTHORS Clark, M. S.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. UK Email:
biohelp@hmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of BAC (pBelobACII) cloned
genomic sequence
The BACs can be obtained from http://www.incyte.com.

FEATURES
source Location/Qualifiers
1..494
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="263K15bD8"
/clone_lib="BAC 263K15"
BASE COUNT 128 a 163 c 45 g 158 t
ORIGIN
Query Match 2.4%; Score 89.6; DB 17; Length 494;
Best Local Similarity 49.08; Pred. No. 5e-12;
Matches 239; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

Qy 1102 ACAACTTCATATGTTGGTGTGACTACTTCTCTATCTGACTAGACTGGCACCAATTTGGTGA 1161
Db 6 ACGACTACTACTACTACTGCGGCTTCTACTACTACTACTACTACTACTACTACTACTACTACT 65

```

```

Qy 1162 ACAGCTACTGTATTTGTTGATGTGCCATATCATACTACCACAACCTGTTACCAGTGAATGG 1221
Db 66 ACGACTACTGTGCTGCTCGACTACTGCGAGGACGACTGCTACTACTGCGGCTGCTTCT 125

Qy 1222 ACAGGAACTACATCACTACCACCACAACCTCGTACCACCACTCAACTGATTCATTTGACACAGTG 1281
Db 126 ACTACTGCTACTACTAGCAGCACTGCTACTACTACTACTACTACTACTACTACTACTACT 185

Qy 1382 GTGCTACAAGTTCCACTGCCAAATCCAACTTTAGTACTACTACTGAATTTGCTCTCAGTCC 1341
Db 186 ACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 245

Qy 1342 TTTCCTAAACCACTACAGTTTACTGCTCTCCAGGTGGTACCGATATCTGTGATTATCAGA 1401
Db 246 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 305

Qy 1402 GAGCACCACAACCACTACTGTCATCTACTACTACTACTACTACTACTACTACTACTACT 1461
Db 306 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 365

Qy 1462 ACTACTGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTTATCAGAGAACCACCAAT 1521
Db 366 ACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 425

Qy 1522 CCAACTGTCACTACAACGAGTATTTGGTCTCAATCTTTGCTACTACTACTACTACTACT 1581
Db 426 ACTGCTGCTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 485

Qy 1582 GCTCCTCC 1589
Db 486 GCTACTAC 493

RESULT 14
LOCUS FR0025683
DEFINITION F. rubripes GSS sequence, clone 154E17aC12, genomic survey sequence.
ACCESSION AL018519
VERSION AL018519.1 GI:2684887
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 450)
AUTHORS Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrana, Y.,
Williams, G., and Brenner, S.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source Location/Qualifiers
1..450
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="154E17aC12"
/clone_lib="cosmid 154E17"
BASE COUNT 96 a 140 c 64 g 144 t
ORIGIN
Query Match 2.2%; Score 82.6; DB 17; Length 450;
Best Local Similarity 49.3%; Pred. No. 3.1e-10;
Matches 208; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 1276 ACAGTGGTGTACAGTTCCACTGCCAAATCCAACTGTTACTACTACTGAATATTTGGTCT 1335

```

```
Db 29 ACTGAGACNTCAAATAACTCAGTGGCTCTTAACCTCTACTACTAATGCTACTACTACT 88
QY 1336 CAGTCCTTTGCTACCACTCAGTACTGCTCTCTCCAGGTGGTACCGACTGTGATT 1395
Db 89 ACTGCTACTGCTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACT 148
QY 1396 ATCAGAGGCCCAACCACTACTGCTACTACTACTACTACTACTACTACTACTACTACT 1455
Db 149 GCTACTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACT 208
QY 1456 ACTACTACTACTACTACTGCTCTCTCAGGTGGTACTGACTCAGTAAATATACAGAACCA 1515
Db 209 ACTGCTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACT 268
QY 1516 CCAATCCAACTGCTACTCAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACA 1575
Db 269 ACTGCTACTGCTGCTACTGCTGCTACTGCTGCTGCTACTGCTACTACTACTACTACT 328
QY 1576 GTTACTGCTCTCCAGGTGGTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 1635
Db 329 GCTACTACTGCTGCTACTACTGCTACTACTACTACTACTACTACTACTACTACTACT 388
QY 1636 GTCAACCACTGAATATTGGTCCCAATCTTAGCCACCACTACTACTGCTACTGCTCT 1695
Db 389 GCTACTACTACTGCTACTGCTGCTGCTGCTACTGCTACTGCTACTGCTACTGCTACT 448
QY - 1696 CC 1697
Db 449 NC 450
```

```
RESULT 15
LOCUS BM181884
DEFINITION fv51b11.y1 Sugano SJD adult male danio rerio cDNA clone 5412044 5'
similar to contains element TARI repetitive element ; , mRNA
sequence.
BM181884
VERSION BM181884.1 GI:17512842
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 641)
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
WasNU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
```

```
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3-Er from Amersham
High quality sequence stop: 527.
Location/Qualifiers
1. .641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
```

FEATURES

source

Search completed: April 29, 2003, 20:16:15
Job time : 3377 secs

```
/tissue_type="whole body"
/dev_host="adult"
/notes="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG);
Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGGCCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed and donated by Dr. Sumio
Sugano (University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGCTGCG and 3' end primer
CGACCTGCGAGCTCGAGCACA."
```

BASE COUNT 218 a 234 c 37 g 152 t
ORIGIN

```
Query Match 2.2%; Score 81.4; DB 13; Length 641;
Best Local Similarity 48.1%; Pred. No. 7.7e-10;
Matches 297; Conservative 0; Mismatches 311; Indels 9; Gaps 2;

QY 2661 ATCTCAATCGTAACTAGTACTGTTCTTACTGCAAGTACAAATGCTGCTGCTACTGCT 2720
Db 34 ATCTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 93
QY 2721 AACTGATGTTAGTGTCTACATCTTCTGATAATGTTTCAAAATCAGAGGTATCAGTTAC 2780
Db 94 TACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 153
QY 2781 AACCGAAACTTCTGTTACAACTATTCAAACTACTTCCAAACCCCAATATCATCTTCCAGT 2840
Db 154 AACCAACAACACTTCTCCAGCTACTACATCACTTCTCCAAACAACAACACTTCTCCAACA 213
QY 2841 ATCATTCAGTCAGTTGCTTCTCAATTCCTCAAGTGTTCAGAAAGTGAAGTAAAGTTACAT 2900
Db 214 AACAACTTCTCCAGCTACTACATCACTTCAACAACAACAACAACAACACT--TCTCCAG 270
QY 2901 TACAAGCAATGGAGACAACAACAAGTGTGCTCATGATTTCACAATCTACTTCCACTGAAT 2960
Db 271 AACTTCAACAACAACAACAACACTTCTCCAGCTACAACTTCTCCAGCAACAACAACAACAT 330
QY 2961 TGAATTTGTAAACACCACTGTTCTACTAAAGTTTACCACCTGTCTGTTTCTTCTAATCTGA 3020
Db 331 AGCTACATCAACTTCTCCAAACAACAACAACACTTCTCCAGCTACATCAACTTCTCCAACA 390
QY 3021 TTTGACTAGTGAACCAACAATAATACAGAGACAACAACCACTACATATATCAACTTCAAA 3080
Db 391 AACAACTTCTCCAGCTACTACAACTTCTCCAAACAACAACAACAACAACACTTCTCCAGCT 450
QY 3081 CTCCATCAGTGAAGATATACCACTATCACCACATCTCAACCTTACAGGTGATATATGGACA 3140
Db 451 TTCAACAAC-----AACAAACAACCTTCTCCAGCTACATCAACTTCAACAACAACAACA 504
QY 3141 ATCAACCAATCCAGTTCTCAACTGTGCAACAAGTACTTTTAGCATCTGCAAGTGAAGAAGA 3200
Db 505 TCAGCTACAACTTCTCCAGCAACAACAACATCAACTTCTCCAGCTACATCAACTTCAACAAC 564
QY 3201 CAACAAAACGGTTCTCATGAATCAGCATCCACAAGTTTGAACCAAGTATGGGTGAAGA 3260
Db 565 AACAAACAACCTTCTCCAGCTACTACATCAACTTCAACAACAACAACAACAACACTTCTCCAG 624
QY 3261 TTCTGGATTAACTACTTT 3277
Db 625 TTCTCCAGCAACAACAT 641
```

